8-11-2015

G-quadruplex Polymorphism: a Winding Road from Structure to Therapeutics

Ekaterina M. Stroeva
Georgia State University, ekaterina.stroeva@yale.edu

G-QUADRUPLEX POLYMORPHISM:

A WINDING ROAD FROM STRUCTURE TO THERAPEUTICS

by

EKATERINA M. STROEVA

Under the Direction of Markus W. Germann, PhD

ABSTRACT

Structural and folding patterns of promoter regions (c-MYC, BCL) and telomeric (TTAGGG repeats) G-Quadruplexes were explored and their interactions with small molecules were tested. EAO-108 and MHI-21, symmetric heptamethine cyanines, were identified to exhibit strong interaction with the telomeric G-Quadruplex (mixed hybrid-1), while promoter region G-Quadruplexes were discriminated against. ZK-26 binds strongly to the c-MYC derivative and it is capable of discriminating between (3+1) hybrid-1 and -2 telomeric G-Quadruplexes with an initial binding preference towards the latter. To aid future binding studies, a novel labeling method is also described in this work.

INDEX WORDS: G-Quadruplex, Telomeric, DNA labeling, Cyanines, Oncogene Promoter

G-QUADRUPLEX POLYMORPHISM:

A WINDING ROAD FROM STRUCTURE TO THERAPEUTICS

by

EKATERINA M. STROEVA

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2015

G-QUADRUPLEX POLYMORPHISM:

A WINDING ROAD FROM STRUCTURE TO THERAPEUTICS

by

EKATERINA M. STROEVA

Committee Chair: Markus W. Germann

Committee: W. David Wilson

Maged Henary

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2015

DEDICATION

To Sarah Vu Nguyen – colleague and friend. Quite selfishly, I hope you will find this thesis helpful, complete and easy to consult. Not so selfishly, I wish you to have as much fun as I did while researching, experimenting, learning about quadruplexes, and growing up.

ACKNOWLEDGEMENTS

There are no words that could precisely convey my gratitude for the opportunity to spend four years in Dr. Germann's lab. And if those words do exist, I still do not know them. As of now, I think those were the best days of my life. I would like to thank Dr. Markus Germann for letting me enjoy my time in the lab, allowing me to grow up at my own pace, and giving me time to figure things out on my own. Thank you for patiently working with me when I needed attention, while trusting me to work independently when I wanted that. Thank you for generously sharing your knowledge and wisdom with me and honestly giving me your opinion when I asked for advice. Although I was not able to grok many things, I truly appreciate those that I did understand. I would like to thank my lab mates for being great colleagues and friends. Dr. Jin Zhang, thank you for introducing me to the lab and for mentoring me along the way. Dr. Alexander Spring, thank you for patiently dealing with my childish behavior and for not giving up believing in my abilities. Shauna, Bello, and Matt, thank you all for being awesome friends. I would like to thank Marina Evich for always cheering me up when I felt down. Thank you for patiently tutoring me on proper NMR techniques, which I should have learned eons ago. Thank you for allowing me to participate in your arginine project – I learned a lot from you. Narinder Kaur, thank you for allowing me to collaborate with you on your two-site NMR project. It was a lot of fun working with you and thank you for also being a very supportive friend. I would also like to thank Schuyler Aebi for briefly helping me on the G-Quadruplex project. Sarah Vu Nguyen, thank you for making the G-Quadruplex project going. Thank you for always being eager to partake in any (sometimes out of the box) experiments and ideas and for encouraging me to keep going when things did not work as planned.

I would also like to thank our collaborators and my thesis committee members Dr. W. David Wilson and Dr. Maged Henary. In particular, I would like to thank Hang Huynh and Sarah Laughlin from Dr. Wilson's group, and Eric Owens from Dr. Henary's group.

I am greatly indebted to Georgia State University and the department of Chemistry for providing me with fantastic undergraduate and graduate opportunities. I would like to thank my parents for supporting me along this wonderful journey and always encouraging me to aim for the best.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

The fact that the Guanine-rich nucleic acid sequences could form stable structures with four strands (now known as G-Quadruplexes) was first mentioned in the work of David R. Davies and Gary Felsenfeld in 1968 [1,2]. However, since the ground breaking discovery in 1953, scientists believed that there was not much else to learn in the DNA structural field, and therefore, other macromolecules, like RNAs and proteins, required more attention.

Work in the DNA structural studies remained dormant for over two decades. In the 1990s, the revolutionary work of Elizabeth Blackburn demonstrated elevated telomerase activity in cancer cells. It was then understood that the stabilization of the G-quadruplex DNA could impede the telomerase enzyme and potentially prevent cancer [3]. In addition, the G-Quadruplex forming sequences were also found in the promoter regions of many oncogenes, which could be silenced if the G-Quadruplexes were stabilized.

A rational design of small molecules, which could specifically target the G-Quadruplex DNA, remains an important topic of research in the biomedical field. However, currently there are still no clinically approved therapeutics available [4].

1.1 Structure

A possible arrangement of a G-tetrad (quartet), four guanine bases connected by hydrogen bonds, was first proposed in 1962 [2]. Now termed a building block of the G-quadruplexes, it consists of four guanine bases connected by Hoogsteen hydrogen bonds as shown in Figure 1.1. There are a total of eight hydrogen bonds: four N1-H hydrogens are hydrogen-bonded with O6 and four N2-H hydrogens are hydrogen-bonded with N7. Hydrogen bond acceptors are O6 and N7 and hydrogen bond donors are NH1 and C2NH. The surface area

of the G-quartet is approximately 135 $Å^2$, which is approximately twice as large as the surface area of the canonical Watson-Crick base pair [4].

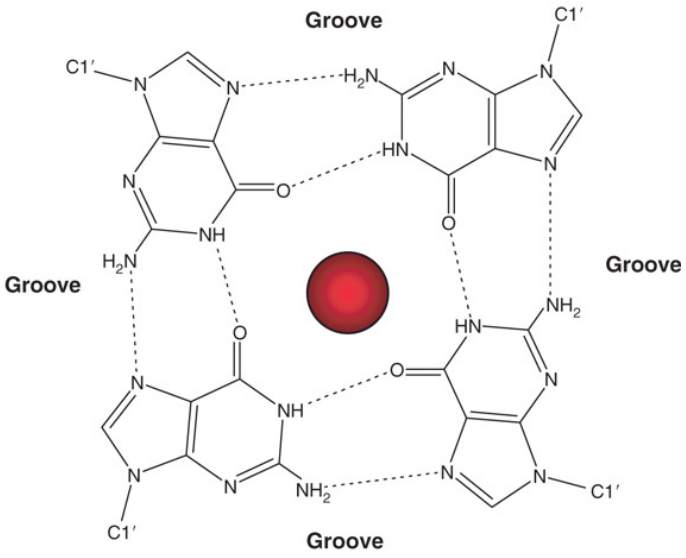


Figure 1.1: Arrangement of the G-quadruplex building block.
Representative figure was adopted from [5]. G-quartet consists of four guanine bases stabilized with eight hydrogen bonds. A cation is schematically shown in red in the center of the quartet. Most common cations are potassium and sodium.

A G-quadruplex molecule consists of several tetrads stacked on top of each other and can arise from one (unimolecular), two (bimolecular) or four (tetramolecular) equivalent DNA strands. All G-quadruplex molecules are stabilized by cations. The most common cations are potassium and sodium. Because of its smaller radius, a sodium cation is in the plane of the quartet while potassium is situated in between the tetrads [4].

Structural polymorphism and commonly used terminology of some unimolecular G-quadruplex topologies are represented in Figure 1.2. Antiparallel G-Quadruplexes contain two strands pointing in one direction, while the other two strands point in the other direction. This kind of arrangement can be supported by differently arranged loops, which connect the strands. Diagonal loop connects two guanines located at the opposite corners of a tetrad. Diagonal loops

always connect two anti-parallel strands (two strands pointing in the opposite directions). Lateral loops connect guanines on the same side of the tetrad. Lateral loops also always connect two anti-parallel strands. Chain reversal (propeller or double-reversal side) loops are required to connect two parallel strands, hence their names. A propeller loop always connects guanines of different tetrads and sits in the groove of the quadruplex.

It is important to note that the G-quadruplex-specific terminology has not been standardized yet. Across different research groups and time periods of the manuscripts, names of loops and structure configurations have often varied. For example, propeller, chain reversal and double-reversal side loop are among the terms used to describe a loop that connects two strands of the same orientation.

Parallel G-Quadruplexes contain three propeller loops which are required to connect four parallel strands. Mixed or (3+1) G-Quadruplex structures have three strands pointing in one direction, while only one strand is antiparallel. Hybrid-1 and hybrid-2 are subtypes of the mixed topology. Both hybrid structures have two lateral loops that connect the antiparallel strands, while the two parallel strands are connected with the chain-reversal (propeller) loop.

G-Quadruplex forming sequences can be conserved to TTAGGG repeats as in the case of human telomeric overhang or be extremely diverse if they are found in the promoter regions of oncogenes. Diverse topology is a feature characteristic of G-Quadruplexes. Especially non-telomeric G-Quadruplexes, which have loops of different size and nucleotide composition, add to the conformational complexity.

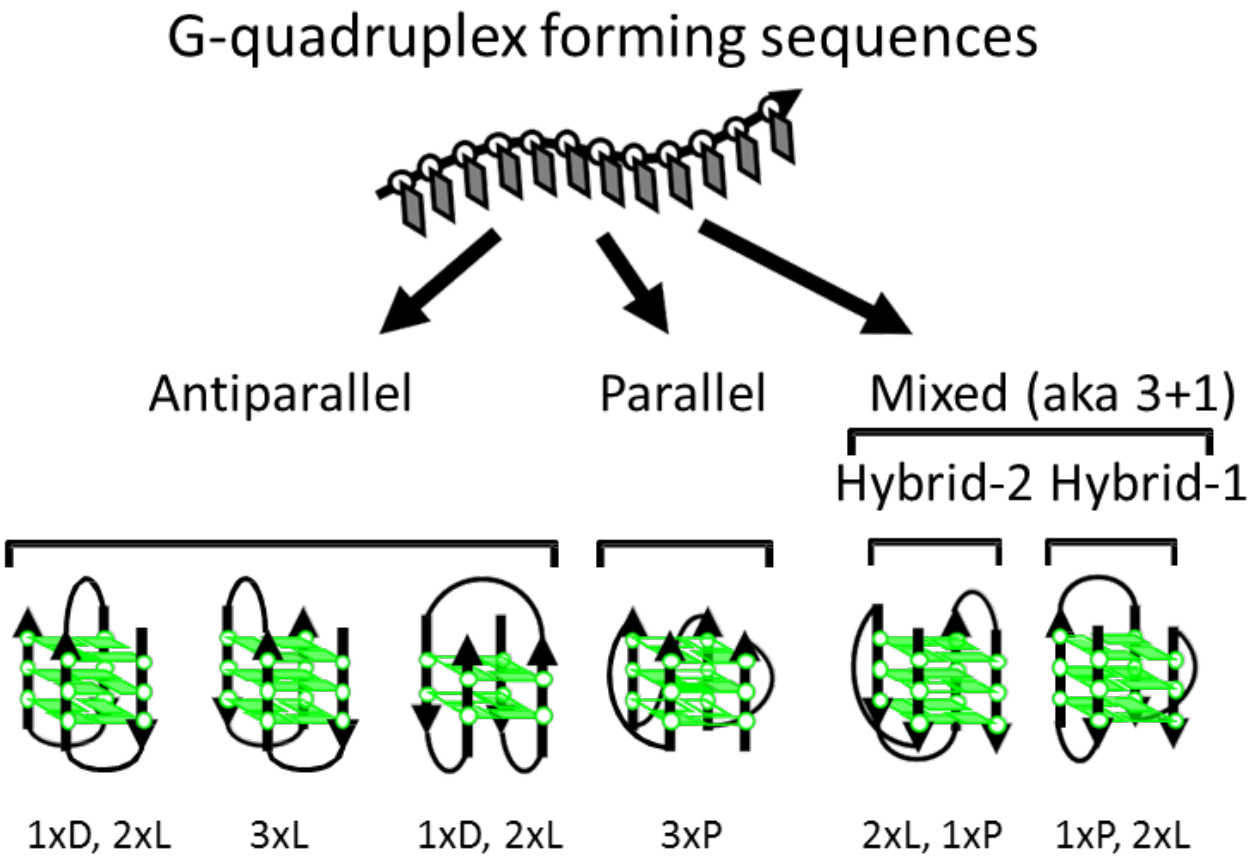


Figure 1.2: Representative human telomeric DNA quadruplex structures and terminology. Loops connecting G-tetrads can adopt different configurations such as diagonal (D), lateral (L), and propeller (P). Figure was augmented with loop properties, modified and adapted from [6]. Specific G-Quadruplex topology described using loop configuration (such as 2xL, 1xP vs. 1xP, 2xL for hybrid-2 vs. hybrid-1) simplifies the analysis [4].

1.1.1 PU-22

The structure of the modified cMYC 22-mer (MYC22-G14T/G23T) (aka PU-22) was published in 2005 by Ambrus et al. [7]. Authors showed that the modified sequence formed a single very stable G-quadruplex structure in potassium solution. Two **G**-to-T mutations at positions 14 and 23 of the wild type MYC22 sequence (TGAGGGTGGG**G**AGGGTGGG**G**AA) were shown to restrict the modified sequence (TGAGGGTGGG**T**AGGGTGGG**T**AA) into one

parallel quadruplex conformation. The melting temperature was determined to be higher than 85^{o}C which suggested a structure of exceptional stability. A 2015 paper by Kim et al. suggests that the exceptional stability of PU-22 is primarily due to its structural features. If compared to the human telomeric G-Quadruplexes, modified c-MYC has a 5-10% larger distance between negatively charged phosphate groups, which leads to weaker charge repulsion [8].

Figure 1.3 represents a schematic model of a parallel PU-22 G-Quadruplex. Three propeller loops connect three guanine tetrads formed from four parallel strands. Three of the four grooves are occupied by the loops. Two propeller loops consist of a single thymine (T) nucleotide, while the other propeller loop consists of T and A. Variable temperature NMR melting studies showed that the single nucleotide loops are more stable than the TA loop. Imino protons corresponding to the guanine bases 13 and 16 (which are connected by a T14A15 loop) are the first to disappear in the imino proton spectrum. In addition, 5'-TGA and 3'-TAA terminal caps were shown to adopt a well-defined conformation and serve as stabilizers. On the contrary, the wild type c-MYC sequence was shown to form a mixture of dynamic conformers, which had been a considerable obstacle for NMR structure work and therapeutics research earlier [7].

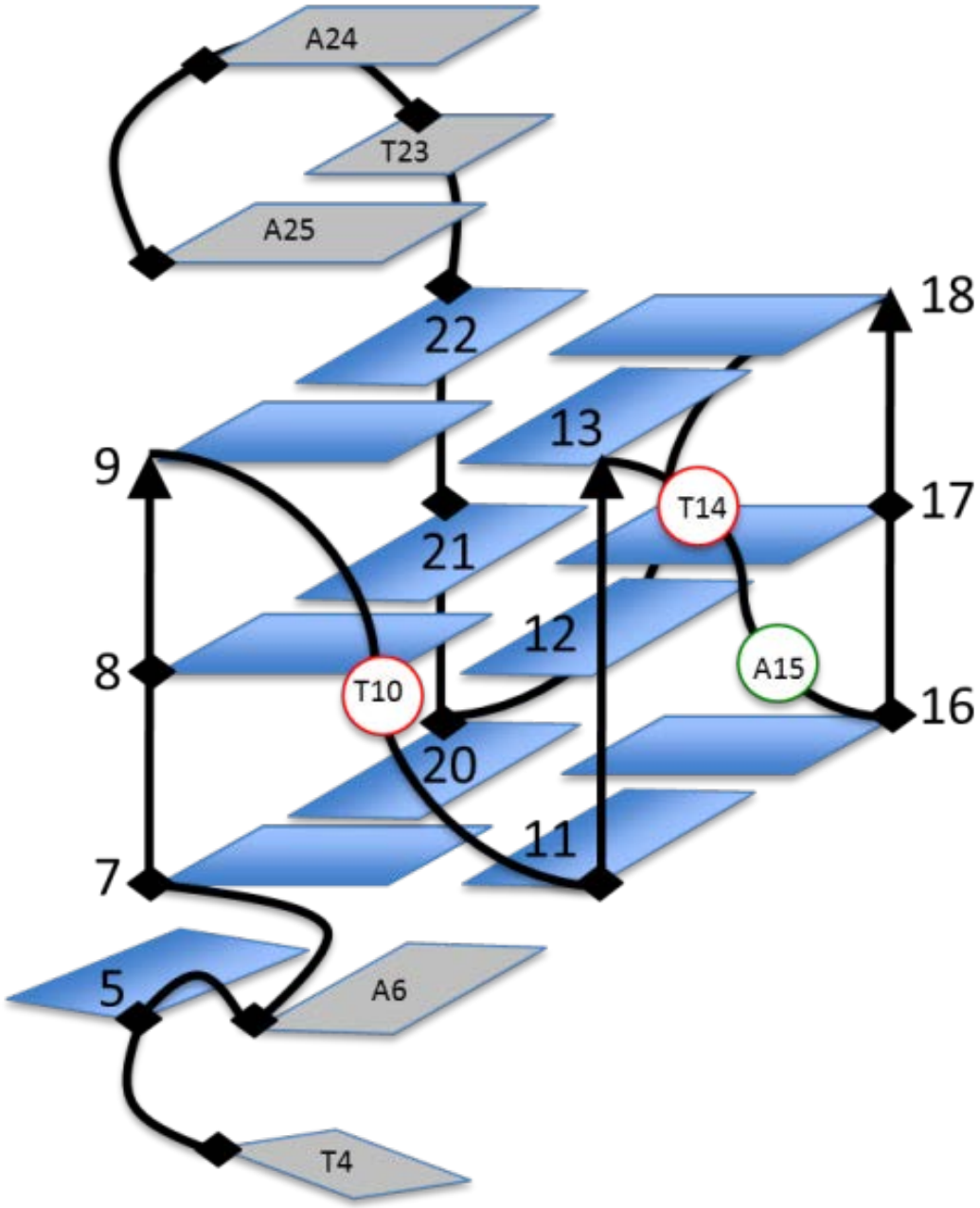


Figure 1.3: Schematic representation of PU-22 G-Quadruplex structure.
Structure is adapted from [7]. G-quaruplex forming sequence is 5'-TGAGGGTGGGTAGGGTGGGTAA-3'. Loops that connect strands of the quadruplex are underlined. Guanine bases are shown as blue rectangles. Guanine at the fifth position is not involved in a tetrad formation while the rest are associated into G-Quartets by Hoogsteen hydrogen bonds. Adenine and thymine loop bases in 5' (TGA) and 3' (AAT) are shown as grey rectangles. Their possible stacking arrangements are represented in the figure. Propeller loop A and T bases are shown as green and red circles, respectively. Strand polarity is shown by arrow pointers. The only unoccupied groove of this structure is on the side pointing away from the viewer (G7, 8, 9 and G20, 21, 22).

1.1.2 BCL-2-MID

The wild type sequence of the BCL2 gene promoter region consists of 39 nucleotides and has been shown to form a mixture of three G-quadruplexes stabilized with potassium cations. A BCL-2-MID sequence is a truncated and mutated 23-mer (GGGCGCGGGAGGAATTGGGCGGG). It was shown to form a stable single antiparallel 1xD, 2xL three-quartet G-quadruplex in the presence of potassium cation represented in Figure 1.4. Compared to the wild-type sequence, this BCL-2-MID is a good molecular model for structural and therapeutic studies of the BCL2 oncogene [9]. A recent 2014 studies identified another highly GC-rich sequence also located upstream of P1 promoter region, and it was shown to be important in BCL2 transcription. The sequence was also shown to form an antiparallel G-quadruplex of the 1xD, 2xL type. However, this intramolecular quadruplex consisted of four G-quartets, which made imino proton NMR spectrum more complex [10].

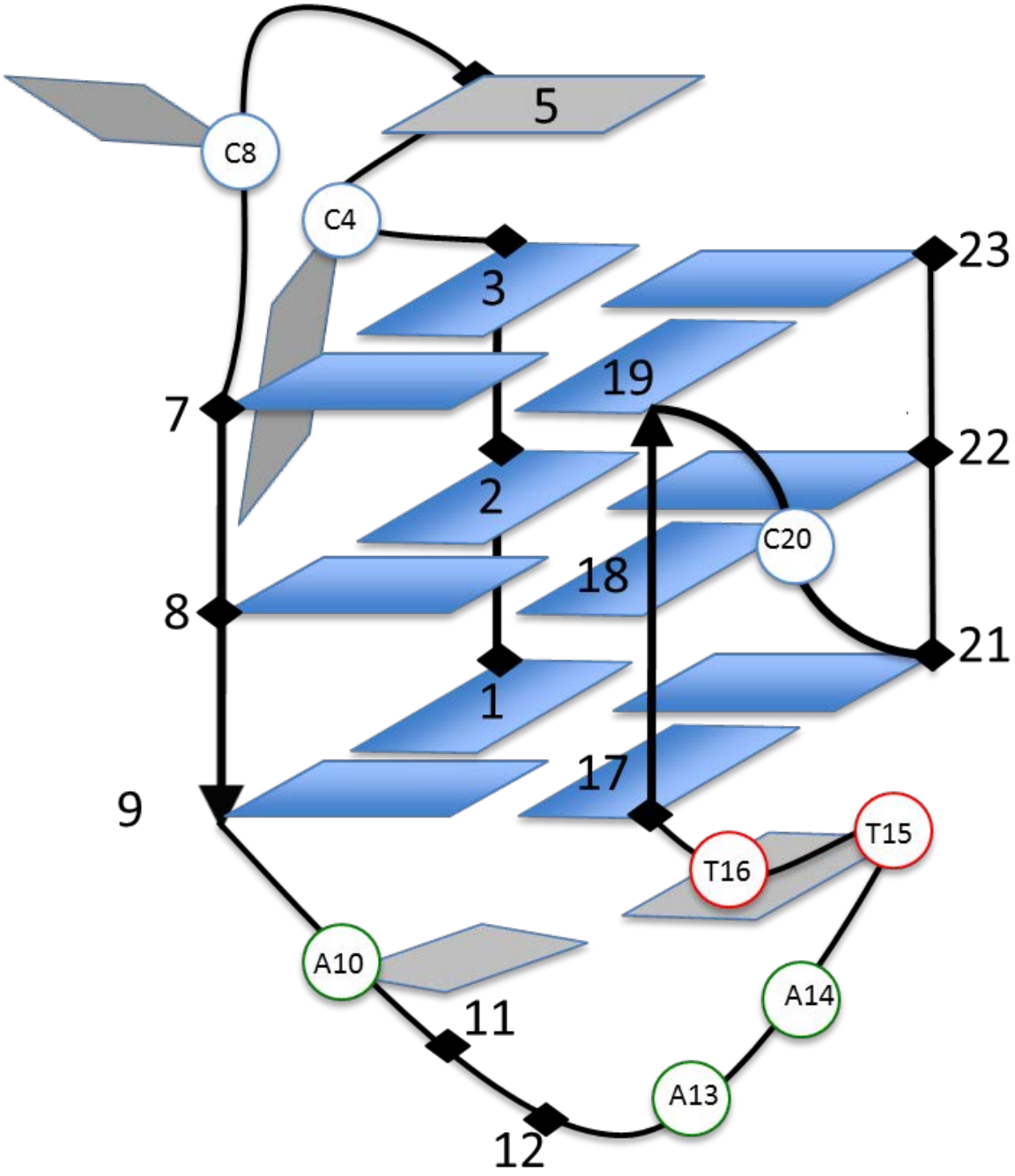


Figure 1.4: Schematic structure of the G-Quadruplex structure formed by the BCL-2-MID sequence. Figure is adapted from Dai et al. [9]. The quadruplex forming sequence is 5'-GGGCGCGGGAGGAATTGGGCGGG-3', where loops connecting quadruplex strands are underlined. Guanine bases associated into G-quartets by means of Hoogsteen hydrogen bonds are shown as blue rectangles. Guanine residues not involved in a tetrad formation are shown as either grey rectangle (G5) or as small black cubes. Adenine, thymine, and cytosine loop bases are shown as either grey rectangles or green/red circles. Orientation of bases represented as rectangles are accurate and demonstrates possible stacking arrangements if any. Strand polarity is shown by arrow pointers. The only propeller loop that occupies one groove of this quadruplex is shown on the side pointing towards the viewer and consists of one C20 nucleotide.

1.1.3 TEL-24

One of the potassium stabilized human telomere repeat G-Quadruplex structures was stabilized by flanking caps (5'-TT and 3'-A) which were added on to a TTAGGG repeats ($(TTGGG(TTAGGG)_3A$). This major mixed (3+1) hybrid-1 form is 95% favored in potassium solution [11].

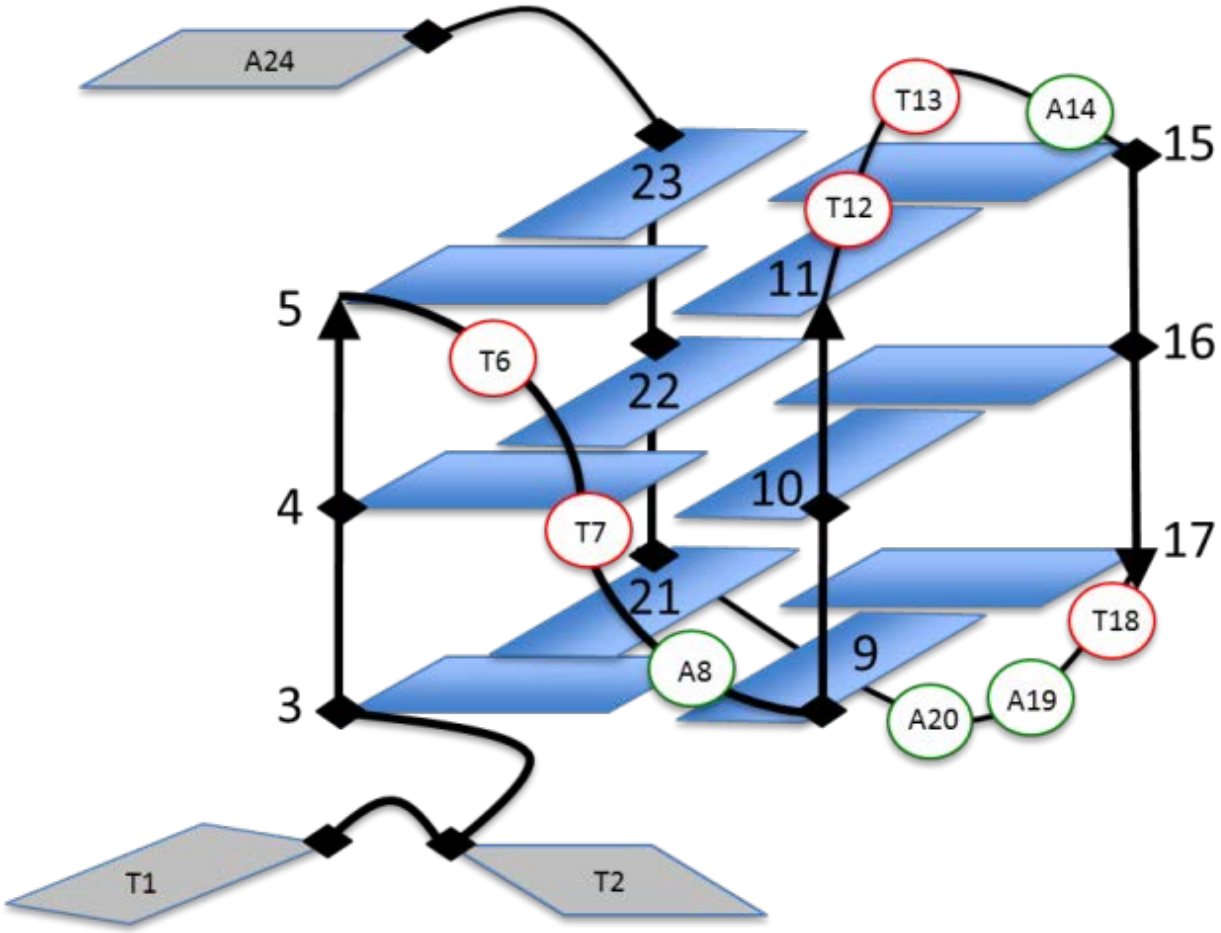


Figure 1.5: Schematic representation of TEL-24 mixed (3+1) hybrid-1 form.
Figure is adapted from [11]. The quadruplex forming sequence is 5'-TTGGGTTAGGGTTAGGGTTAGGGA-3', where loops connecting quadruplex strands are underlined. Guanine bases associated into G-quartets by means of Hoogsteen hydrogen bonds are shown as blue rectangles. All loops that connect guanines consist of TTA, where adenine and thymine bases are shown as green and red circles, respectively. Capping loop residues (3' A24 and 5' TT) are shown as grey rectangles. Strand polarity is shown by arrow pointers. The only propeller loop that occupies one groove of this structure is shown on the side pointing towards the viewer and consists of TTA nucleotides.

The minor form that comprises the other 5% in solution presented a significant challenge to characterize. Currently, equal amounts (50/50) of both TEL-24 hybrid-1 (major) and hybrid-2 (minor) forms can be induced. However, we were not able to induce more than 50% of the minor hybrid-2 form. Different approaches have been developed to induce 50/50 ratio of the major/minor form. One of the methods is presented in the results section 3.1.4, while the other was published recently. A paper published by Bessi et al. in June 2015 described another induction procedure (KCl injection). Using the high field NMR and expensive labeling techniques, allowed the group to confirm the structure of the "minor" form, mixed (3+1) hybrid-2. Kinetic and thermodynamic pathways were characterized and hypothetical structures of intermediates were proposed in the article [12].

Both hybrid-1 and hybrid-2 G-Quadruplex structures feature three strands running in the same direction, which requires the presence of at least one propeller (chain-reversal) loop. The other two loops are lateral loops. Geometrically, the main difference between two hybrid structures is the order in which the loops follow. For TEL-24 sequence read in the 5' to 3' direction, the hybrid-1 structure has a propeller loop, which is followed by two lateral loops. To the contrary, the hybrid-2 structure has two lateral loops, which are followed by the propeller loop (Figure 1.5 and Figure 1.6).

In addition, it is important to point out that the long telomeric TTAGGG repeats form tandem quadruplex repeats (bead-on-a-string model) connected by TTA loops. However, the precise mode of connection and interaction between individual G-quadruplexes is still unknown and difficult to elucidate from techniques that study individual single quadruplexes in solution or crystal [4,13].

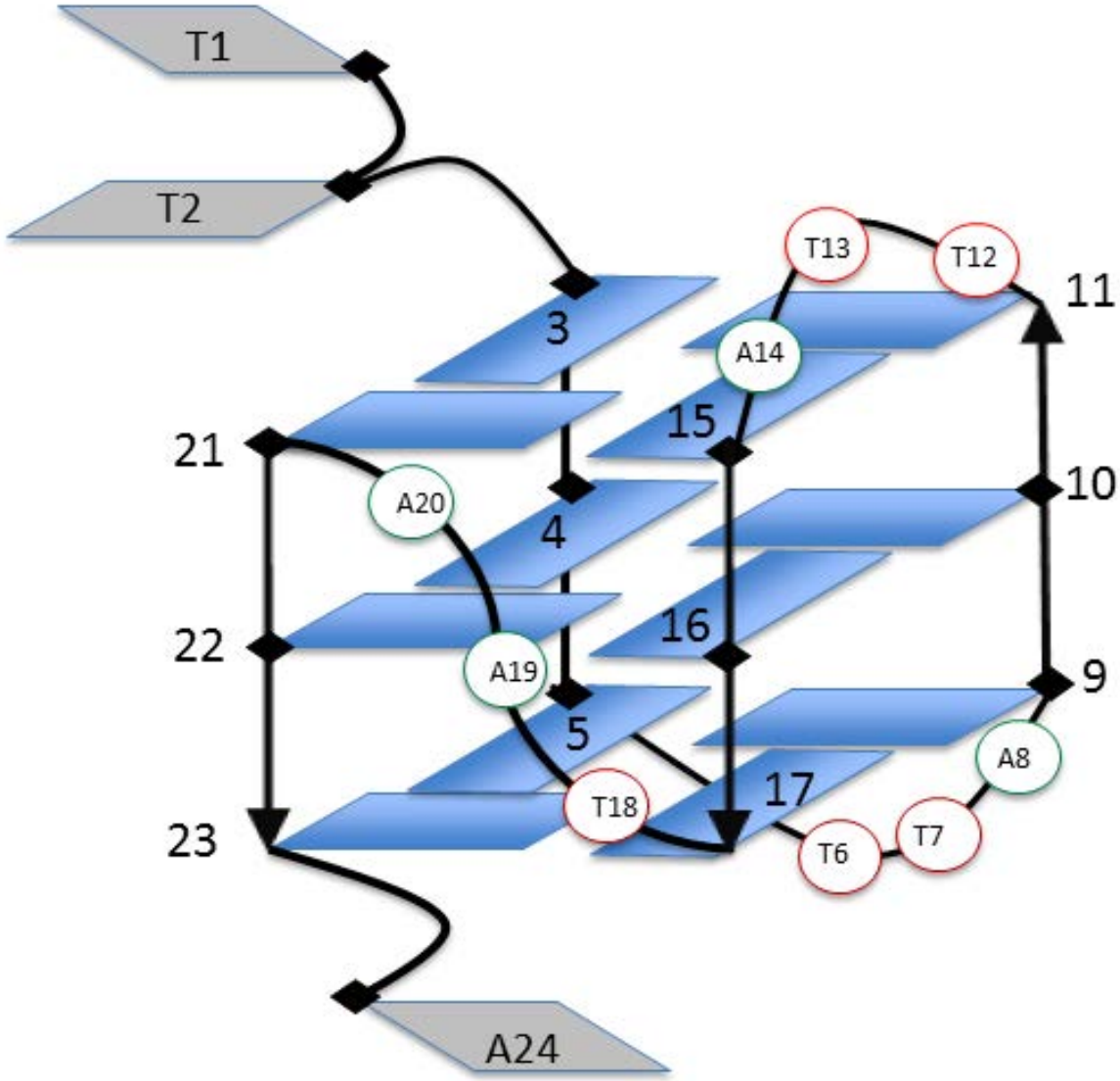


Figure 1.6: "Minor" form of TEL-24. Mixed (3+1) hybrid-2 G-Quadruplex.
Figure is adapted from [12]. The quadruplex forming sequence is 5'-TTGGGTTAGGGTTAGGGTTAGGGA-3', where loops connecting quadruplex strands are underlined. Guanine bases associated into G-quartets by means of Hoogsteen hydrogen bonds are shown as blue rectangles. All loops that connect guanines consist of TTA, where adenine and thymine bases are shown as green and red circles, respectively. Capping loop residues (3' A24 and 5' TT) are shown as grey rectangles. Strand polarity is shown by arrow pointers. The only propeller loop that occupies one groove of this structure is shown on the side pointing towards the viewer and consists of TTA.

1.2 Importance

1.2.1 Telomeric Sequences

A large body of research on telomeric G-quadruplex structures was compiled by Stephen Neidle [4]. Telomers are structures located at the ends of chromosomes, which consist of G-rich DNA sequences combined with different telomeric proteins. Human telomeric DNA is relatively large (3-15 kb). Two complementary strands, which consist of TTAGGG and AATCCC repeats, form a duplex DNA. In addition, a single-stranded overhang is located at the 3'-end of the chromosome. The DNA sequence in that region is capable of folding into a string of G-quadruplexes when various telometic proteins are not associated with it.

Cellular apoptosis process is triggered when the telomeric DNA has been significantly shortened after several replication cycles, because the DNA polymerase cannot efficiently replicate blunt ends. However, telomeric repeats can be added back onto the DNA sequence by enzyme telomerase, which usually is not present in human somatic cells. The presence of the telomerase in a cell makes it immortal, which was shown to have a significant (80-85%) implication in human cancers (osteosarcoma cells use Alternate Lengthening Pathway to maintain lengths of their telomeres). Therapeutics specifically targeting the enzyme telomerase have been under development for over two decades.

1.2.2 Non-telomeric Sequences

Sequences in the promoter region of c-myc capable of forming G-Quadruplexes were the first studied among the non-telomeric sequences [4]. Overexpression of c-MYC oncogene has been linked to a variety of human cancers. The transcription of the oncogene is primarily controlled by NHE III_1, a nuclease hypersensitive element III_1 of the c-MYC promoter, which is located upstream from the P1 promoter of the oncogene (Figure 1.7). The coding C-rich strands

of that element can adopt i-motif conformations, while the non-coding G-rich strand can adopt G-quadruplex conformations, which repress transcription of the c-MYC [7,8].

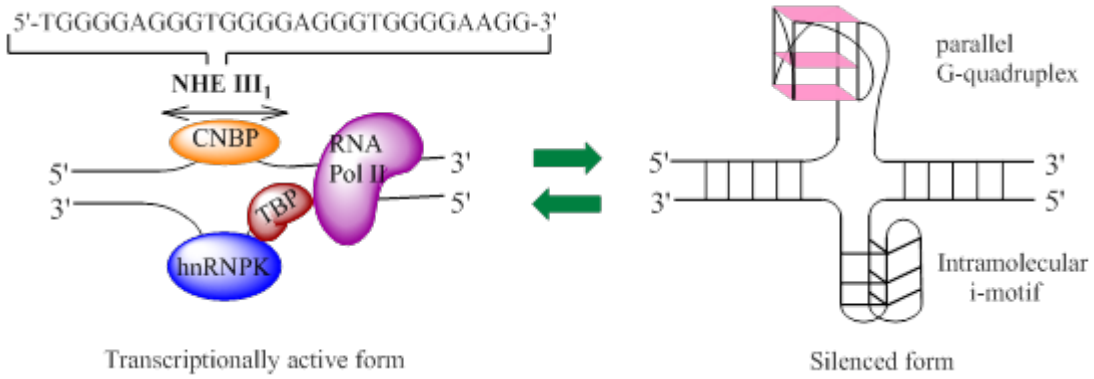


Figure 1.7: Transcription regulation of c-MYC oncogene with G-Quadruplex formation.
Figure is borrowed from [14]. The transcriptionally active complex of NHE III_1 is shown with RNA Pol II representing RNA polymerase reading the genomic information. Transcriptionally silenced form is shown to form parallel G-Quadruplex on one-strand and i-motif on the complementary strand.

Transcription of the BCL2 (B-cell CLL/lymphoma 2) gene is also down regulated by G-quadruplex formation. The BCL2 gene produces a protein which is a part of the mitochondrial membrane. The protein was found to be an inhibitor of cell apoptosis and its overexpression was found in a variety of cancers, including B-cell and T-cell lymphomas. A GC-rich P1 promoter region and the region upstream of it were shown to regulate transcription of the BCL2 gene. [9,10,15].

1.3 Targeting with drugs

The transcription of a gene can be stalled by formation of a G-quadruplex molecule. Therefore, stabilization of the G-quadruplex by small molecules will prevent RNA polymerase from transcribing the gene. Since the promoter region quadruplexes are generally not stacked on top of each other and are flanked with single-stranded loops (which connect them to the rest of the double stranded DNA), intercalation between the tetrads, groove binding, as well as top and bottom stacking of potential small molecules presents a plausible way to stabilize the quadruplexes [4].

On the other hand, telomeric G-quadruplexes exist in tandem, so top and bottom stacking of small molecules may stabilize several G-quaruplexes at the same time and efficiently inhibit the telomerase activity.

1.4 NMR

Nuclear Magnetic Resonance (NMR) spectroscopy is one of the two methods used to determine bio-macromolecular structure at high resolution. Unlike X-ray crystallography, which needs a suitable crystal, NMR allows for the monitoring of the interconversion of multiple structural forms present in solution. In addition, thermodynamic and kinetic properties of the system can be studied by varying concentration, ion presence, temperature, and solvents [16]. Drug titration experiments shown in the results section were performed by tracking chemical shift changes of imino protons. Aromatic and sugar proton chemical shifts can supplement imino proton information.

For all quadruplex structures used in this study the ^{1}H assignments were published. Imino and aromatic proton spectra were always compared to the published spectra. Detection of

phosphodiester bonds in a DNA molecule provides crucial data for the structure of a DNA backbone. In general despite their obvious relevance ^{31}P spectra of nucleic acids, in particular G-Quadruplexes are rarely recorded. This is in part due to the availability of the required probe and the presumption of extensive resonance overlap. Only BCL-2-MID G-Quadruplex had a ^{31}P proton-decoupled spectrum published (

Figure 1.8) [9].

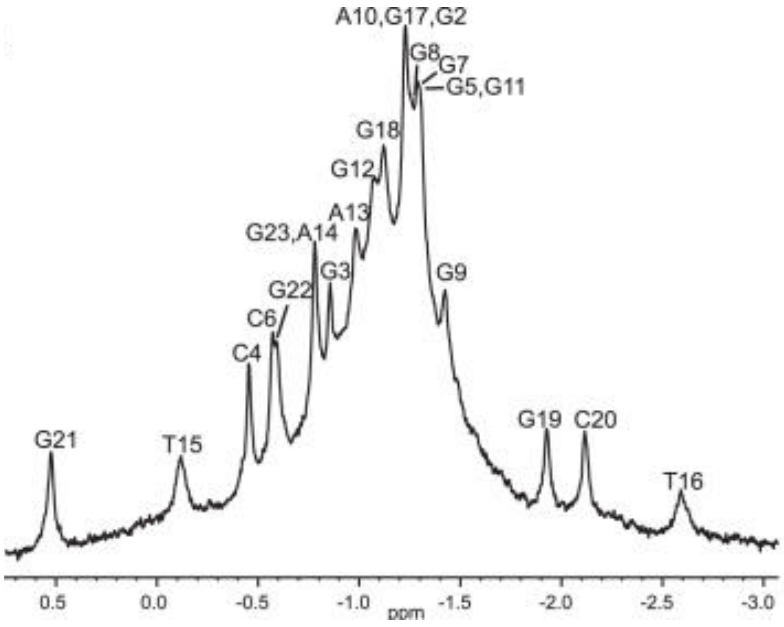


Figure 1.8: Phosphorus spectrum of BCL-2-MID phosphate backbone.
1D proton decoupled ^{31}P NMR spectrum was adopted from [9]. Phosphorus assignment accounts for every phosphodiester present in the BCL-2-MID G-Quadruplex.

1.5 H+-Exchange Method

A novel method which allows for selective silencing of AH8 and GH8 NMR signals has been developed in our lab and is currently under refinement. The aromatic region of the quadruplex DNA sequence rich in guanine repeats can be significantly simplified. Very

importantly, this method does not require expensive isotope labeling and can be carried out within a period of two weeks while only requiring regular DNA molecules as starting material. Guanine and to a lesser extent adenine exhibit a slow exchange of the protons attached to the carbon in position 8 shown in Figure 1.9. Exchange rates for these protons were measured in 1976 by tritium incorporation [17]. The exchange mechanism of GH8 and AH8 involves donation of the proton to the base, which suggests that the rates of exchange can be altered with high pH, DNA strand concentration, and temperature [16].

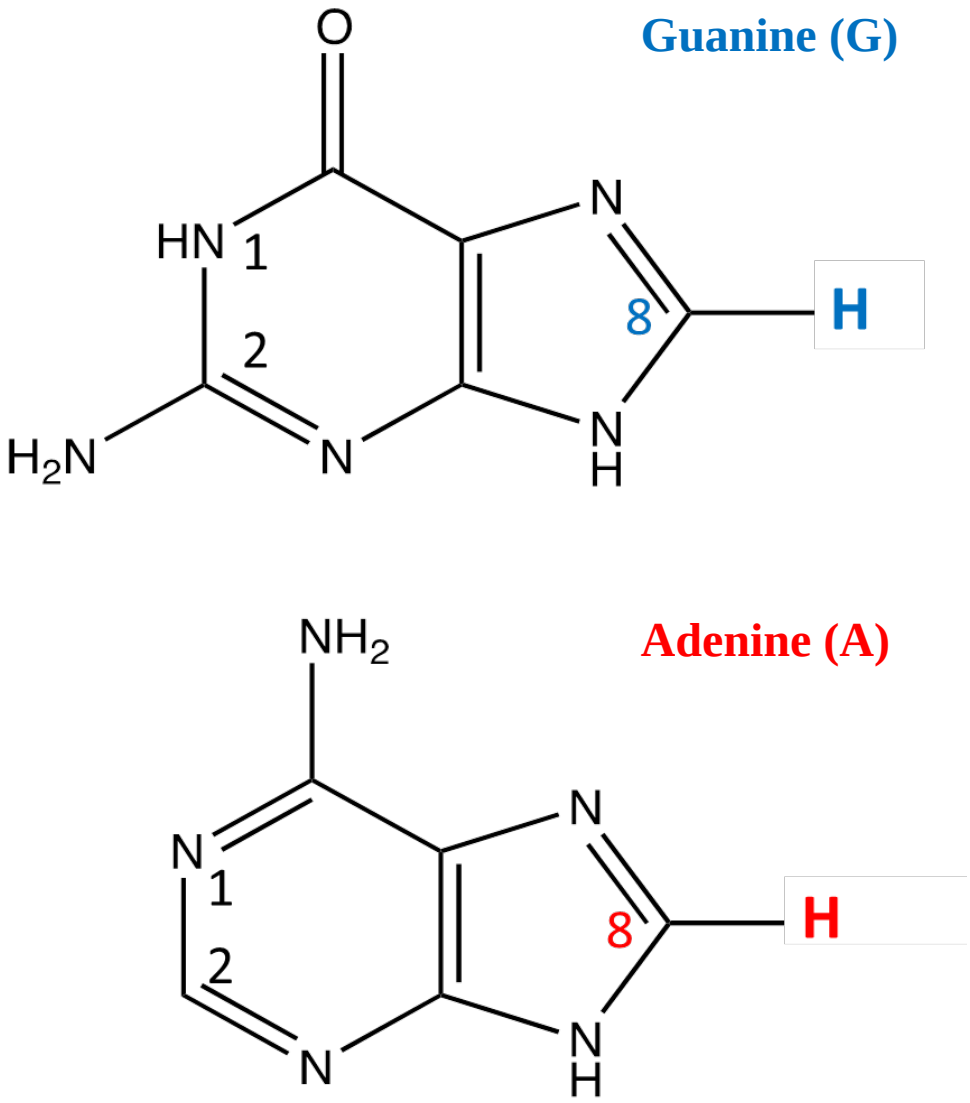


Figure 1.9: The purine coding bases in DNA and RNA.
Guanine (**G**) is 2-amino-6-keto purine and adenine (**A**) is 6-amino purine. G has an imino group at N1. Both guanine and adenine are connected at the sugar moiety at N9 in the imidazole ring. GH8 and AH8 are slowly exchangeable. Protons attached to a carbon in the 8th position resonate in the aromatics region of the NMR spectrum. Due to their abundance in the G-quadruplex molecules, silencing of their signals tremendously simplifies aromatic region of the NMR spectrum [16].

2 MATERIALS AND METHODS

This section outlines general guidelines for sample preparation and data acquisition. Any deviations to this procedure are described in the results section of the thesis.

2.1 DNA Samples

The DNA nucleotides: PU-22 d[TG(AGGGTGGGT)$_2$AA], BCL-2-MID d[GGGCGCGGGAGGAATTGGGCGGG], TEL-24 d[TT(GGGTTA)$_3$GGGA], TEL-22 d[AGGG[TTAGGG]$_3$], and duplex control RI-10-mer d(GCGAATTCGC)$_2$ were purchased from Integrated DNA Technologies (Coralville, IA, USA). The concentration was calculated using the absorbance at 260 nm from the manufacturer provided extinction coefficient. Stock solutions of the oligonucleotides were prepared in the doubly deionized water.

Titration NMR samples of DNA were prepared in 90 % H_2O/10% D_2O, 10-20 mM TRIS (Tris(hydroxymethyl)aminomethane) or potassium phosphate buffer and 20-50 mM potassium chloride. DSS (4,4-dimethyl-4-silapentane-1-sulphonic acid) was added as an internal reference. EDTA (ethylenediaminetetraacetic acid) was added as a chelating agent for potential sample contamination with paramagnetic metals. Sample pH was adjusted to 6-7.

2.2 Small Molecules

Small molecules used in this study were synthesized by the research team of Dr. Maged Henary at GSU.

2.3 NMR Experiments

NMR experiments were performed on an Avance 600 spectrometer, equipped with a 5 mm QXI $^{1}H\{^{31}P, ^{13}C, ^{15}N\}$ Bruker probe. Imino proton spectra (1024 scans for 100 uM DNA sample) were acquired at 35 ^{o}C using the 1-1 jump and return technique [18] and WATERGATE w5 [19] pulse sequence. Sweep width was adjusted to 24 ppm.

3 RESULTS AND DISCUSSION

3.1 G-quadruplex structures used in this study

Exchangeable imino protons are trapped in the hydrogen bonds formed between nucleic acid base pairs. They resonate approximately in the 16 to 11 ppm region of the 1D ^{1}H NMR spectra of DNA sequences. The unique chemical shifts of imino protons make them a convenient monitor to probe for structural perturbations. Imino proton regions of PU-22 [7], BCL-2-MID[15], and TEL-24 [11], quadruplex forming DNA sequences, were previously assigned and published by different research groups.

The aromatics region of a ^{1}H NMR spectrum is significantly more crowded and is not as readily available to monitor the structural perturbations. However, a novel proton-exchange method proposed in section 3.5 allows to efficiently silence GH8 and AH8 signals, which can tremendously simplify ^{1}H NMR spectra, especially that of G-Quadruplex DNA.

In addition to being the most abundant and sensitive proton (^{1}H) nuclei to detect, ^{31}P nucleus is also used to monitor the phosphate backbone of the DNA. Assignment of the 1D proton-decoupled ^{31}P NMR spectrum of BCL-2-MID was also previously published by the same group [15].

3.1.1 Imino protons

G-quadruplex forming sequences found in the oncogene promoter regions were commercially synthesized and 1D ^{1}H NMR spectra were recorded. Using previously published assignments, all NMR signals in the imino proton regions were correlated with their particular imino protons.

Figure 3.1 shows imino proton spectrum of PU-22 recorded on a 600 MHz NMR instrument. Optimal conditions for NMR work were found by varying pH, salt, and temperature

conditions. This beautifully resolved spectrum of a parallel G-Quadruplex formed with the PU-22 sequence was recorded at pH 6 and a temperature of 35 ^{o}C. The presence of a potassium counter ion is important for quadruplex formation, and the correct annealing procedure must be followed to obtain one predominant conformation (as described in sections 3.1.3).

The middle tetrad of this G-Quadruplex consists of guanine bases with their imino protons (GH1) trapped in the hydrogen bonds between the two neighboring bases: G8, G21, G17, and G12. The 3'-end containing tetrad can be monitored with imino protons 9, 22, 18, and 13, while 7, 20, 16, and 11 resonances correspond to the 5'-end containing tetrad.

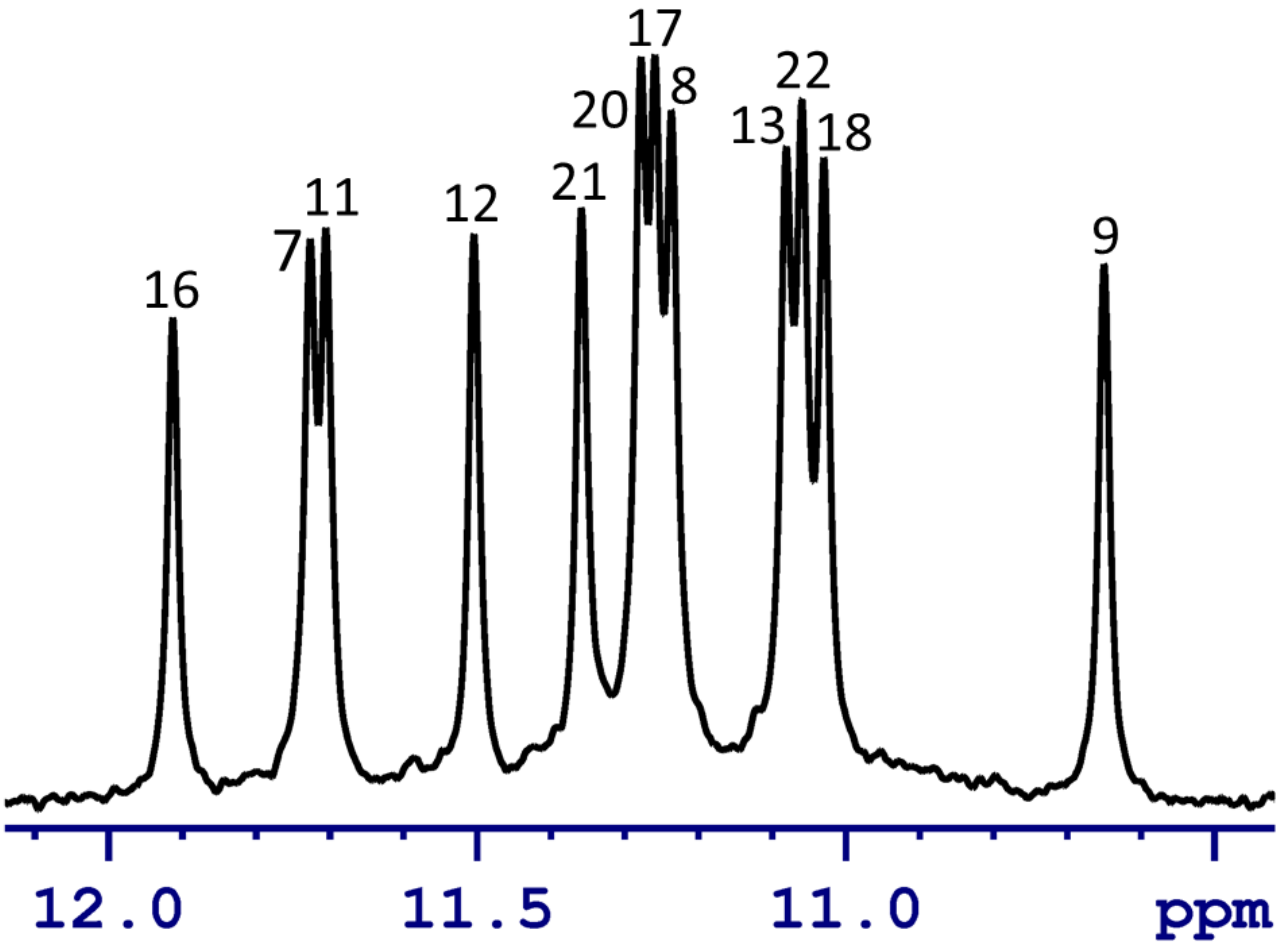


Figure 3.1: Imino proton region of PU-22 ^{1}H NMR spectrum.
Each peak corresponds to one imino proton of the guanine base. Assignment was published in 2005 by Ambrus et al. [7].

The BCL-2-MID quadruplex forming sequence was also probed for the most suitable NMR conditions and spectrum shown in Figure 3.2 was recorded at 35 ^{o}C and pH 7. The same annealing procedure must also be used to ensure formation of the G-Quadruplex structure. Weak signals in the imino proton region suggest that more than one form is present in solution, but a major conformation corresponds to the one published previously [15]. Each major NMR signal was associated with a particular imino proton. The middle tetrad of the quadruplex can be monitored with signals labeled as 8, 2, 22, and 18. Intriguingly huge, the seven-nucleotide lateral loop consisting of A10, G11, G12, A13, A14, T15 and T16 connects guanines 9 and 17, which are part of the 5'-end containing tetrad. Dai et al. showed that A10 and T15 stack to the G-tetrad and a reversed Watson-Crick hydrogen bond was detected at a low temperature (TH3). G3, G7, G19, and G23 imino protons correspond to the 3'-end containing tetrad [9].

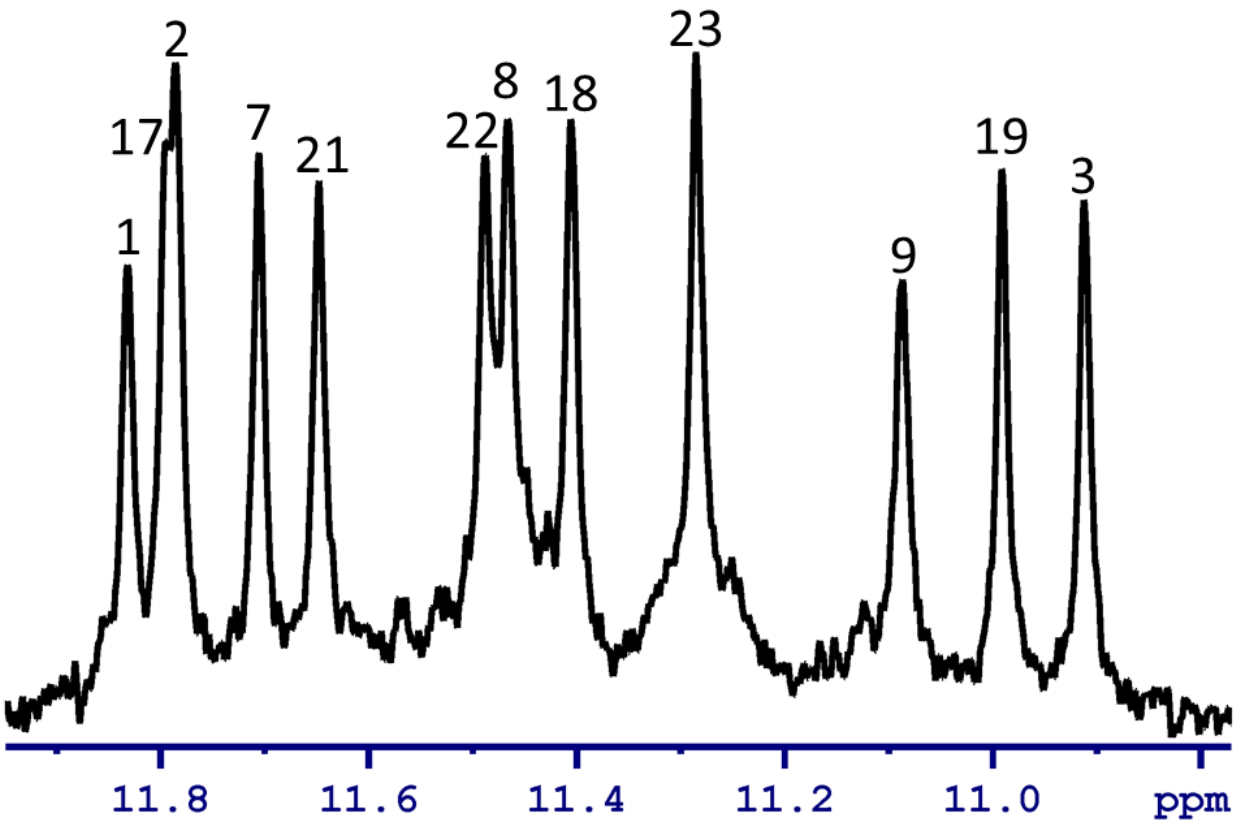


Figure 3.2: Imino proton region of BCL-2-MID ^{1}H NMR spectrum.
Imino protons of guanine bases are assigned based on published data [15].

Among other biologically important G-quadruplex forming sequences are those found in human cells in the telomeric ends. A single-stranded DNA overhang consists of up to 30 TTAGGG repeats. The TEL-22 sequence is a wild-type telomeric DNA sequence capable of folding into a mixture of two monomeric topologically different G-quadruplex structures under physiologically relevant conditions [20]. In the presence of potassium ions in solution, TEL-22 was shown to fold into a mixture of different G-quadruplex topologies.

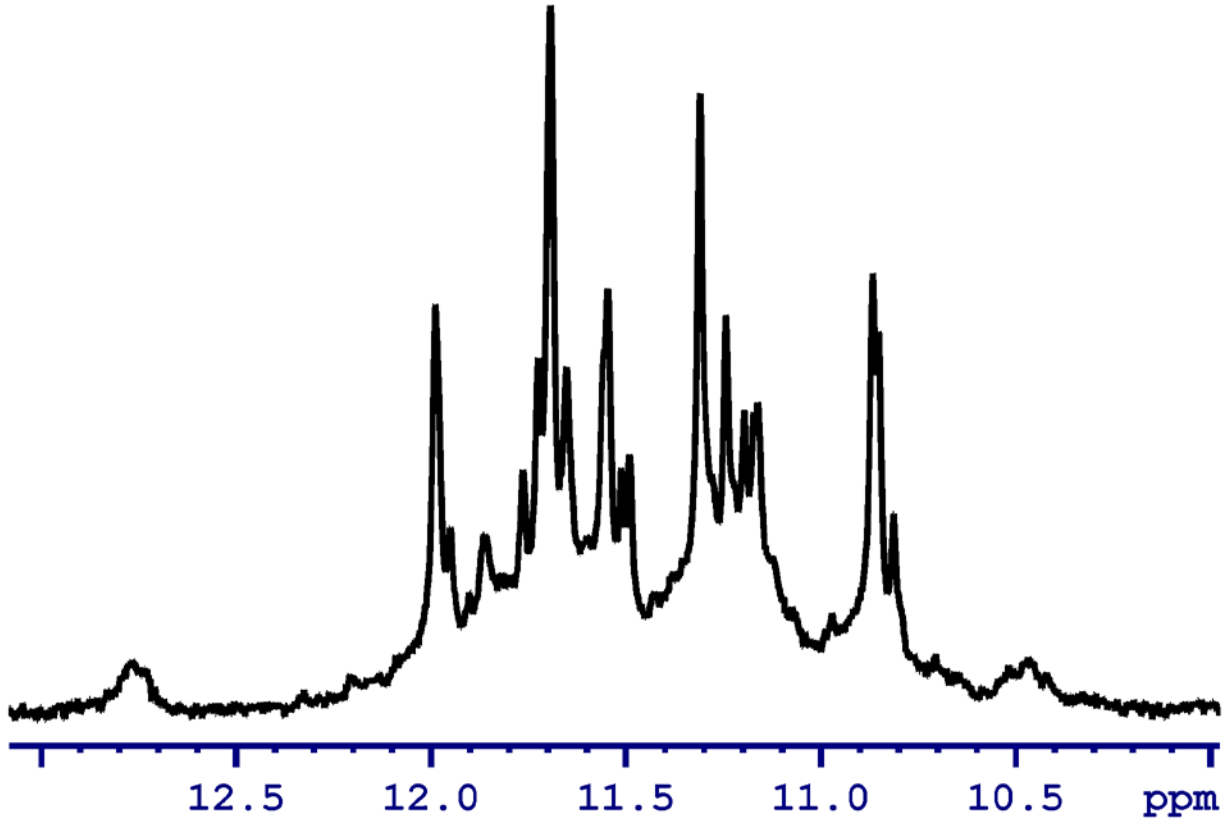


Figure 3.3: Imino proton region of TEL-22 ^{1}H NMR spectrum.
This spectrum was obtained after the wild-type TEL-22 sequence was annealed at 85ºC and allowed to cool down over time. A mixture of structures is present in solution. Previous studies indicate that crowding, variable solvent conditions, and most importantly, loop nucleotide mutations are capable of inducing a single conformation over the mixture of structures [4].

Certain modifications of the 3'- and 5'- ends of the TEL22 sequence were shown to stabilize one (3+1) conformation over the other [20]. For example, TEL-24 (5'-TT; 3'-A) can

fold into a thermodynamically stable hybrid-1 form. A specific procedure described in 13.1.4 has to be followed to achieve a confomationally pure sample. Spectrum of the TEL-24 sequence folded into a hybrid-1 form is shown in Figure 3.4.

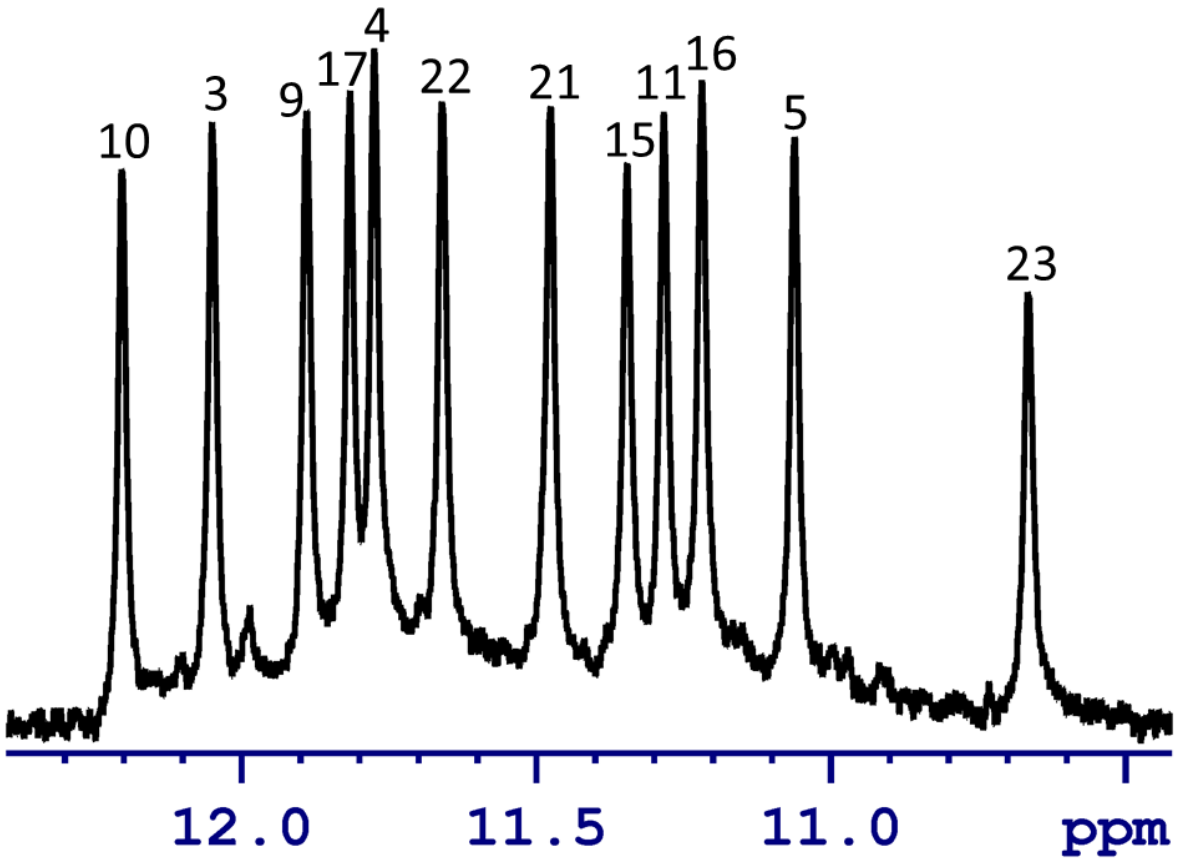


Figure 3.4: Imino proton region of TEL-24 ^{1}H NMR spectrum.
Each peak corresponds to one imino proton of the guanine base. Assignment was previously published in 2006 by Luu et al. [11]

3.1.2 Phosphorus

A DNA backbone can be directly monitored with ^{31}P NMR spectroscopy. To our advantage, the PU-22 spectrum, which is shown in Figure 3.5, accounts for every phosphodiester in the molecule. Structural perturbations of the quadruplex can thus be easily monitored. In addition, Figure 3.6 shows a well resolved ^{31}P NMR spectrum of TEL-24.

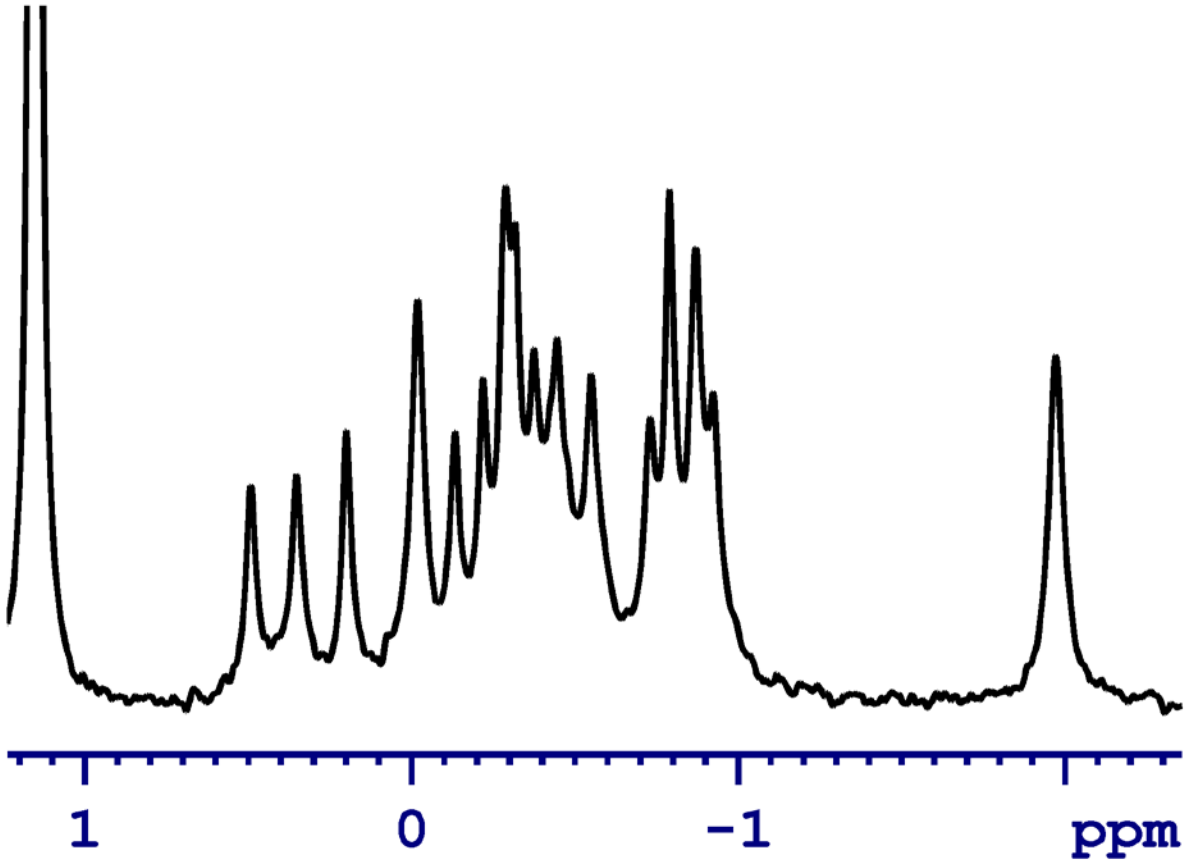


Figure 3.5: Proton-decoupled ^{31}P spectrum of PU-22.

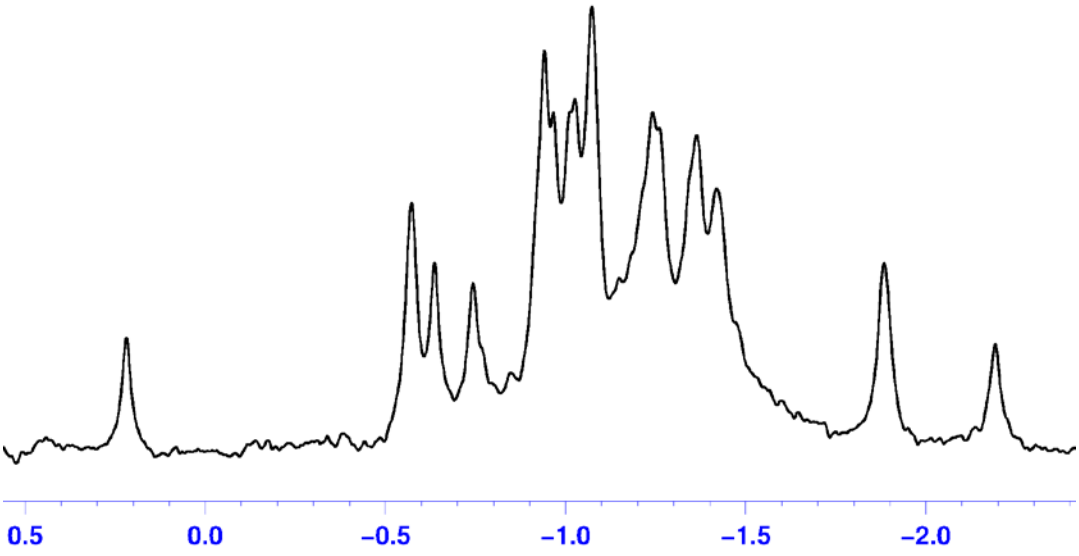


Figure 3.6: Proton-decoupled ^{31}P spectrum of TEL-24.

3.1.3 Multiple forms of PU-22 and BCL-2-MID

To efficiently elucidate a single conformation from a mixture of partially-folded and folded quadruplex topologies, intricate annealing procedures are sometimes created. Procedures described below were successfully used for the specified sequences in this work, but they might not be the most efficient methods.

A major form of both PU-22 and BCL-2-MID samples can be induced under similar conditions. A DNA powder, which was previously desalted with KCl, is pre-mixed with the buffer, solvent, EDTA and DSS, but the salt is not added. The mixture is heated to 85 ^{o}C for approximately 10 minutes. While the solution is still hot, the necessary amount of KCl is added. The sample is then allowed to cool down to a room temperature slowly over time.

After the above procedure was applied to both PU-22 and BCL-2-MID sequences, imino proton spectra were recorded to confirm the presence of the intended conformation as a major form in solution.

3.1.4 Multiple forms of TEL-24

The procedure described in section 3.1.3 was developed independently, but it is similar to a common annealing procedure used to form major conformations of G-quadruplexes. It is also similar to the KCl injection method used by Irene Bessi et al. published very recently, which was used to study the TEL-24 conformers [12]. However, a different procedure to induce a 50/50 ratio of hybrid-1/hybrid-2 confirmations of TEL-24 was independently developed in our lab (described below).

It is important to note that unlike our pH-denaturation/renaturation method, the KCl injection method can be applied to a sample of TEL-24 only once. After the solution of KCl was

injected into the NMR tube, the sample would not be readily available for a second induction procedure. According to the published method, the sample will have to be desalted and purified overnight [12].

A pH-denaturation/renaturation method of induction calls for about 10 minutes of preparation time on average. It is comparable to the 8 minutes reported as necessary to complete induction by the KCl injection method.

A sample of potassium-desalted TEL-24 sequence is mixed with the desired amount of KCl (20-50 mM used in this study) and is heated to 42 – 45 ^{o}C. The sample is then denatured with enough KOH to ensure unfolding, which is confirmed by NMR imino proton spectrum. We have observed that the final pH values in a range of 12-13 will suffice. It is important to note that this method does not degrade the DNA to an extent which will render it unusable. As long as the DNA is not kept at high pH long (we have not tried longer than 15 minutes), the method will work. Figure 3.7 illustrates an imino proton region of the NMR spectrum of the denatured TEL-24. When pH is greater than the pK_a of dissociation of the imino proton, which participates in the base-pairing, the base gets ionized and the hydrogen bond pattern is no longer formed. The imino proton of a guanine base has a pK_a of 9.2 at 25^{o}C and ionic strength of 0.1 [16]. No imino proton signals are observed because the guanine bases are not connected through Hoogsteen hydrogen bonds at that pH. However, currently, we do not know whether this unfolded state is similar to the hypothetical semi-unfolded intermediate states reported by Bessi et al. [12] and to what extent this unfolded state is topologically different from any of the known folded TEL-24 structures.

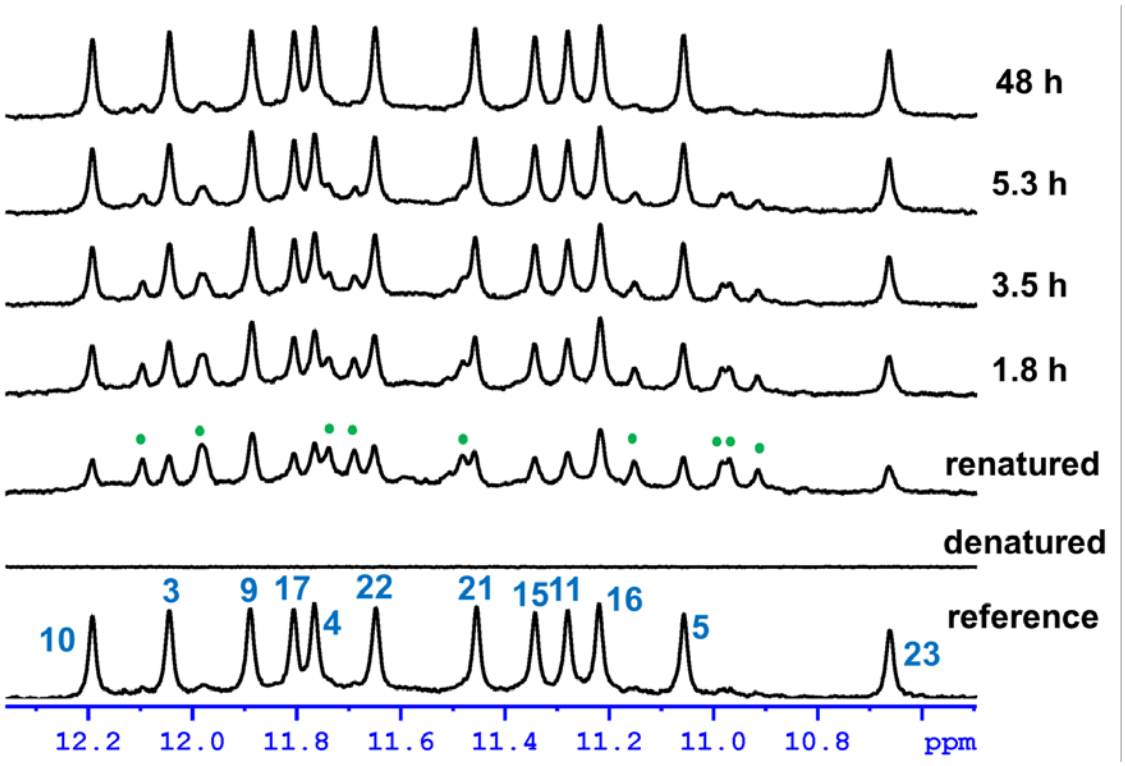


Figure 3.7: Illustration of the minor form formation, disappearance of the minor form (at T=298 K) Reference: major form is labeled according to a previously published assignment corresponding to a 3+1 hybrid-1 topology shown in Figure 1.5. The **reference** spectrum was recorded at T=298 K and pH=7. The **denatured** spectrum (recorded at T=298 K) shows no imino proton signals, because there is no "base-pairing" present at pH=13. The **renatured** spectrum displays both major and minor form, concurrently, where the newly formed minor form is labeled with green circles (assignment of the minor hybrid-2 form was published very recently in [12]). The **renatured** spectrum was acquired at pH adjusted back to neutral and T=298 K. Disappearance of the minor form can be monitored over time. After sufficient amount of time, only the major form remains visible.

After the TEL-24 is denatured, a desired buffer is added to the solution (TRIS-HCl) and pH is adjusted to approximately 6-7.5. The rest of the NMR sample components are added at this time (EDTA, DSS, D_2O/H_2O). Then, the sample is transferred to a preheated (to the same temperature of 42-45 ^oC) NMR tube. The NMR measurements can then be started immediately after the necessary NMR parameters are adjusted. Figure 3.7 illustrates formation of a 50/50 mixture of major/minor form in the renatured spectrum. Since the renaturation procedure together with the NMR set up time take about 10 mins, the time points of the experiment are adjusted accordingly. The first renatured spectrum is in fact recorded 10 minutes after the

induction. More spectra recorded over equal time intervals allow for convenient monitoring of the disappearance of the minor form and appearance of the major form.

A conformationally pure major hybrid-1 form remains in solution after the minor hybrid-2 form gets spontaneously converted over time (Figure 3.8). As illustrated in Figure 3.7, a spectrum of a pure hybrid-1 G-quadruplex is acquired after 48 hours at 298 K.

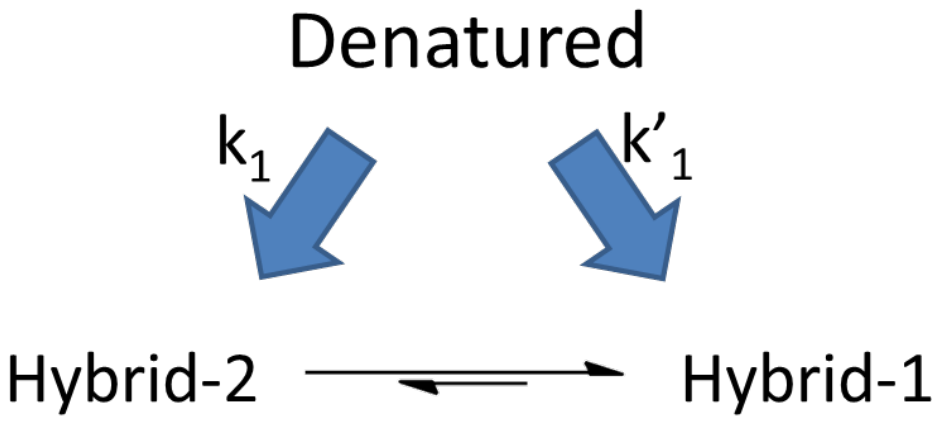


Figure 3.8: Schematic representation of conversion between TEL-24 conformations.

Several pitfalls need to be avoided for a successful induction. The pH-denaturation step, which calls for observation of lack of imino proton signals by NMR, can be skipped to save time; however, if the denaturation was not successful, the renatured form will not be exhibit a 50/50 ratio and significantly less than 50% of the minor form will be present, e.g. 20/80. In addition, if the sample is kept on the heating block for too long or if the temperature of the heating block is set higher than 45^{o}C, a 50/50 ratio will not be formed. For example, if the temperature is set to 50 ^{o}C during the renaturing step, a 40/60 ratio of the minor/major form is observed. In addition, we observed that TEL-24 refolds extremely slowly at temperatures lower than 288 K.

Most importantly, a 50/50 ratio can be induced multiple times in the same sample, with only the total volume and potassium ion concentration change after each induction.

After the NMR sample is prepared, the NMR experiments can be recorded at different temperatures, which allows for kinetic measurements described in section 3.1.5. DNA samples of variable concentrations allowed us to show that the minor-to-major form conversion is independent of concentration in the range studied, which is expected for hybrid-2/hybrid-1 mixtures (described in section 3.1.6).

Physiologically relevant molecular crowding conditions are characterized by the restricted amount of solvent surrounding the macromolecule and by the excluded volume phenomenon. Volume occupied by one molecule becomes unavailable to another molecule. Steric repulsions created in a cell environment, can be mimicked with high concentration of either non-reactive crowding agent (e.g. PEG) or the macromolecules studied. Concentration of macromolecules in a cell environment is reported to be anywhere from 80 to 400 mg/mL [21]. If a particular macromolecular conformation does not change when the concentration of the macromolecule is elevated, then the current conformation could be considered physiologically relevant. Concentration dependence of up to 1.6 mM or approximately 10 mg/ml was tested for TEL-24. Even though it is at least eight times lower than the cell environment, we can suggest that both forms of TEL-24 are potentially valid under physiological conditions in the cell.

3.1.5 Temperature dependence of TEL-24

Kinetic measurements and half-life values were determined from the temperature dependence experiments shown in Figure 3.9.

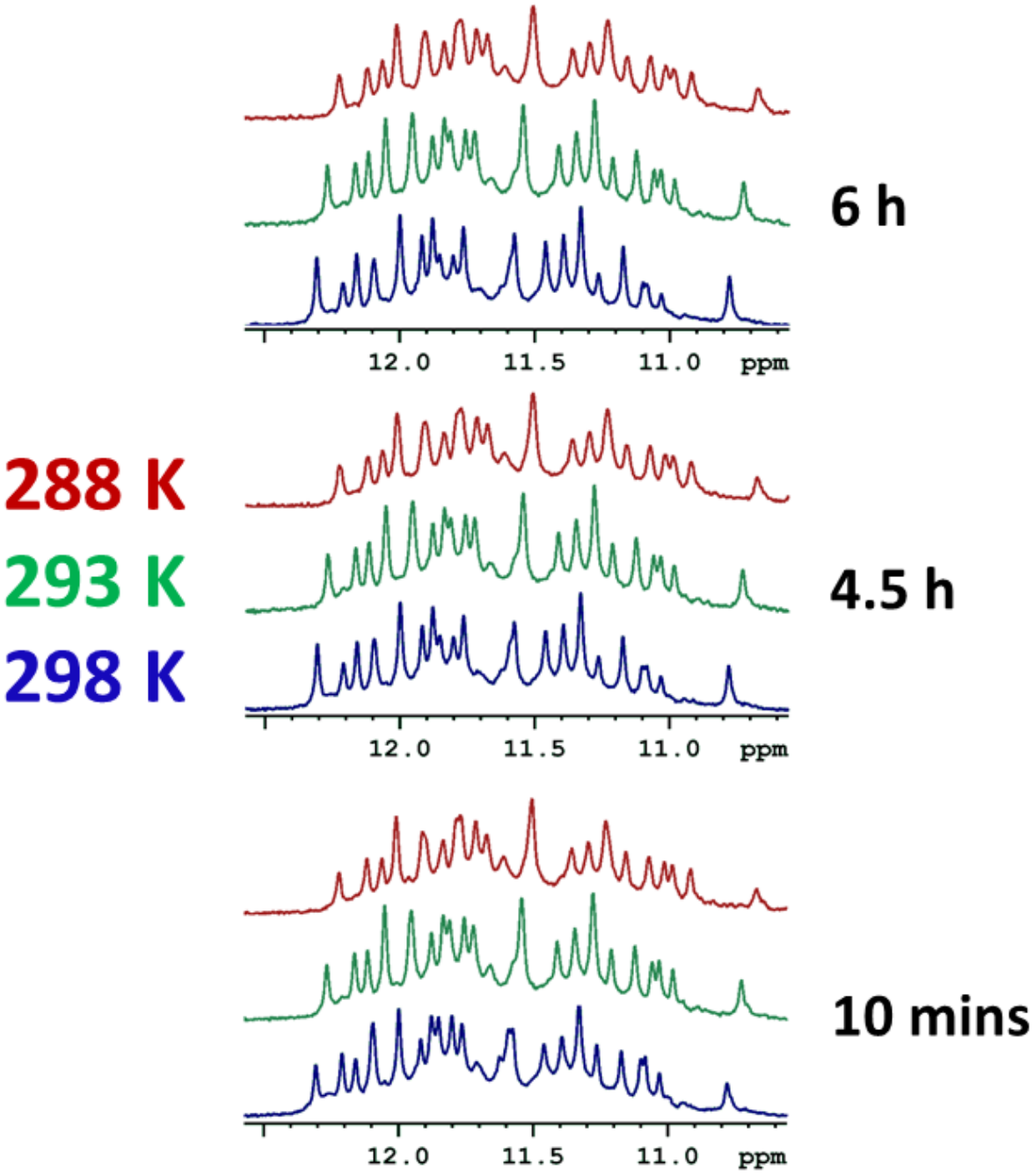


Figure 3.9: Representative NMR data of kinetic measurements at 288, 293, 298 K.
Three data points (10 mins, 4.5 hours and 6 hours after minor form induction) are shown for processes monitored at 298 K (blue), 293 K (green) and 288 K (red).

After the 50/50 induction, the intensities of the minor form peaks (corresponding to the imino protons locked in the hydrogen bonds of the hybrid-2 folded guadruplex) gradually

decrease, while the intensities of the imino protons corresponding to the major form (hybrid-1) increase over time. Intensities of individual hybrid-1 and hybrid-2 peaks were scrupulously measured and recorded. Intensities of all decreasing peaks were added together. Their sum was divided by the sum of the intensities of all increasing peaks. The natural logarithm of the ratio was taken and plotted against time. The equilibrium constant was obtained from the slope of the resulting best fit line, and the half-life in hours was calculated. The process was assumed to follow the first order kinetics. The time that the minor form peaks take to decrease to half of their initial intensity (recorded 10 minutes after induction) is the half-life value rounded to the nearest hour. It is reported in Table 3.1.

Table 3.1: Temperature dependence of half-life (in hours) of the minor-major form conversion.

Temperature (K)	Half-life (hour)
288.0	34
293.0	21
298.0	9

Based on the observed behavior, the minor-major conversion process must have a significant kinetic barrier. Activation Energy (E_a), 23 kcal/mol, was obtained from the slope of the best fit line (ln(k) vs. 1/T). In addition, kinetic data reported by Bessi et al. suggests an activation barrier of 33 kcal/mol [12] (the value is not reported in the paper. It was calculated from the equilibrium constant k_{2r}, which corresponds to the conversion of the hybrid-2 form into the unfolded intermediate reported in the paper). This difference is within the experimental error range.

3.1.6 Concentration dependence of TEL-24

Measurements were also taken for DNA samples with variable concentration. Data is summarized in Table 3.2.

Table 3.2: Concentration dependence of TEL-24 minor form half-life.

Concentration (mM)	**0.4**	**0.8**	**1.6**
Half-life (hours)	26	21	21

Even though the half-life of the minor form in a 0.4 mM TEL-24 sample is slightly different from the 1.8 and 1.6 mM samples, we are inclined to conclude that the kinetic processes are independent of concentration. NMR is not a sensitive technique and resolution of the NMR spectrum is dependent on the concentration. At 0.4 mM total concentration of DNA, imino proton signals of each form were only 0.2 mM in concentration. This concentration is low enough to allow for discrepancies in the measurements. Notably, a common way to record a good NMR spectrum of a low concentration sample is to acquire more scans; however, it is not possible in the timed kinetic experiments.

3.1.7 Kinetics of TEL-24 and Development of Therapeutics

Both hybrid-1 and hybrid-2 forms of TEL-24 are regarded as potential targets for therapeutics development. A 50/50 mixture of two forms presents a convenient system to monitor selectivity of the drugs. However, if hybrid-2 imino protons are monitored, it is important to be aware of the rate of their conversion into the hybrid-1 form. Binding experiments monitored for longer than nine hours at 25 ^{o}C might render the data inconclusive. Successful selectivity studies are reported in section 3.4.

3.2 Targeting of PU-22 with cyanine dyes

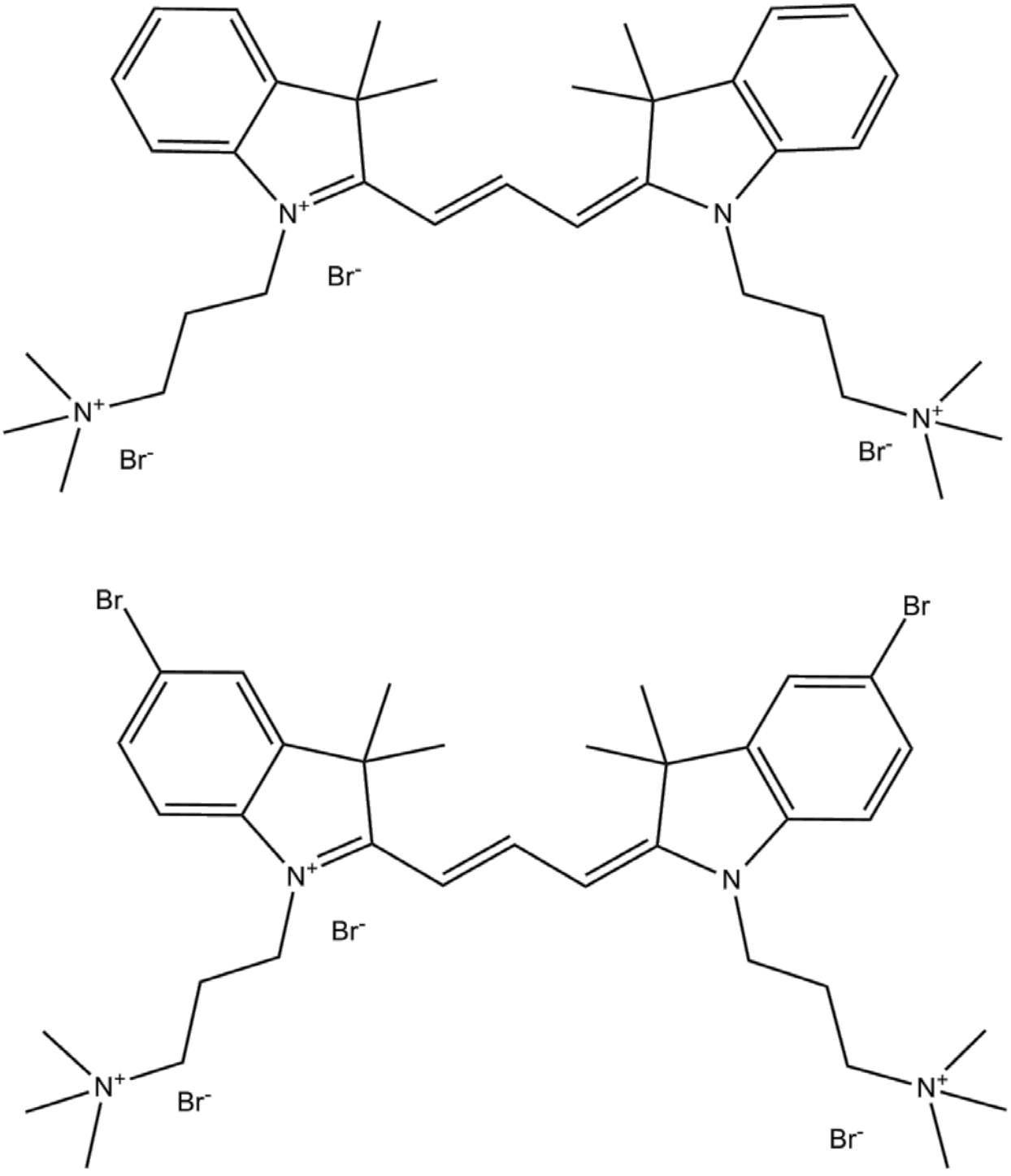


Figure 3-10: Cyanine drugs MH-4 (top) and MH-5 (bottom).
Both compounds are symmetric trimethine cyanines. Both compounds have one positive charge on the cyanine system and most have charged alkyl amine substituents. MH-4 compound does not carry two large Br substituents on the six membered rings.

Binding of the MH-4 and MH-5 cyanine drugs to the PU-22 G-Quadruplex was monitored using perturbations of the imino proton signals of PU-22. As shown in Figure 3.3 and described in section 3.1.1, each imino proton signal corresponds to a particular tetrad. Figure 3.11 illustrates titration data PU-22 with MH-4 (a) and MH-5 (b). The drugs were added to a constant amount of PU-22 at mole ratios of 0:1, 1:1, 2:1, 3:1, 4:1 and 6:1. The bottom spectrum shows imino protons of PU-22 with no drugs added.

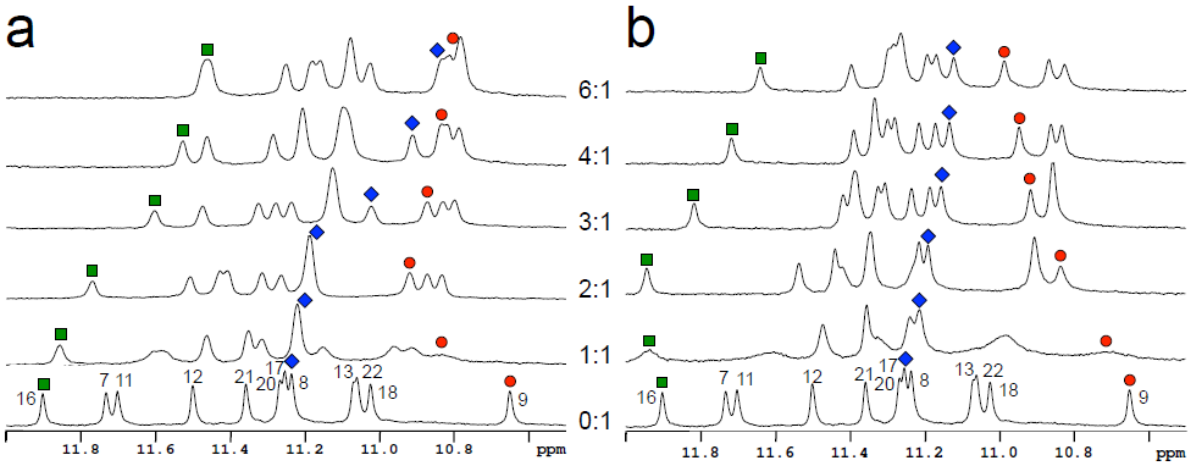


Figure 3.11: Imino proton spectra of MYC22 quadruplex titrated with the parent cyanine 26 (a) and the brominated analogue 31 (b).
Ligands were added to the quadruplex at the ratio indicated on the plot. Selected imino protons of 3'-end (9), middle (8), and 5'-end (16) tetrads are marked.

Based on the imino proton shifts on titration, the primary binding site for compound MH-4 is the 3′-end tetrad of PU-22 (G9, G13, G18, and G22; Figure 3.12 (a)). A secondary binding site at the 5′-end tetrad (G7, G11, G16, and G20) is evident after the 3′-end approaches saturation. Remarkably, no precipitation was observed even at a 10-fold excess of either compound. With the exception of G8 imino proton, the middle tetrad is not affected by the binding on either terminal tetrad. However, it is noted that the shift of G8 imino proton parallels that of G16 imino proton which monitors binding to the 5′-end tetrad. This suggests that the

binding to the 5′-end tetrad causes the chemical shift change of G8 imino proton. A similar binding sequence (first 3′-end and then 5′-end) is also observed for compound MH-5 (Figure 3.12 (b)). Interestingly, the presence of the bulkier bromine substituent results in a much smaller effect on the central tetrad, as shown for G8 imino proton. This is in marked contrast to the result obtained for compound MH-4 and indicates that there are differences in the binding orientation of MH-4 and MH-5 (on the 5′-end tetrad).

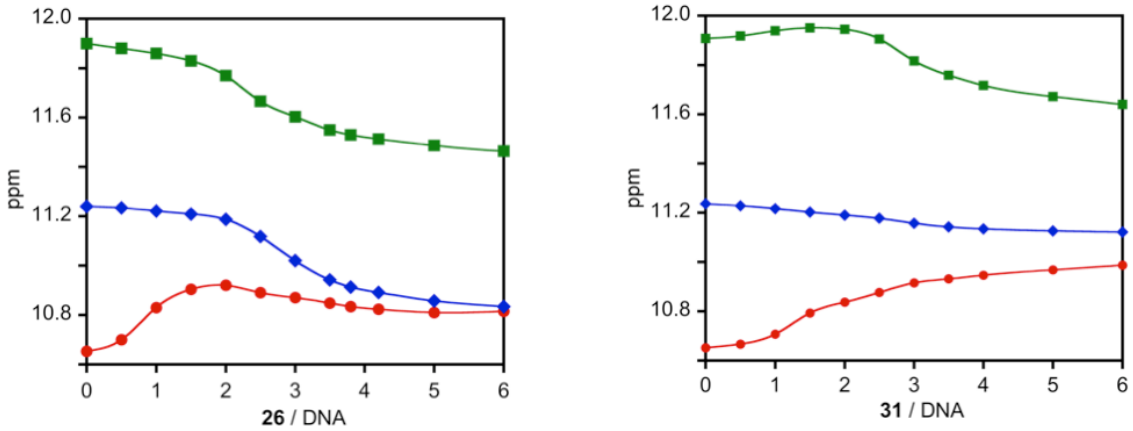


Figure 3.12: Imino proton titration curves of 3'-end tetrad (9), middle (8), and 5'-end (16) tetrad for the parent cyanine 26 (a) and the brominated analogue 31 (b).
The primary binding site for 26 the 3'-end tetrad; a secondary binding site is at the 5'-end tetrad is evident after the 3'-end approaches saturation.

The above data was published as a part of the 2013 paper [22]. Detailed NMR studies of the proposed groove binding mode of MH-5 with other quadruplexes are currently under investigation.

3.3 Discrimination between Different G-Quadruplexes

Similar to the MH-4 and MH-5 drug pair, EAO-108 and MHI-21 are only different by two chlorine residues attached to the six-membered rings of EAO-108 (Figure 3.13). Contrary to the large bromine groups present in the MH-5 molecule, chlorine residues are considerably smaller in size.

3.3.1 EAO-108 and MHI-21

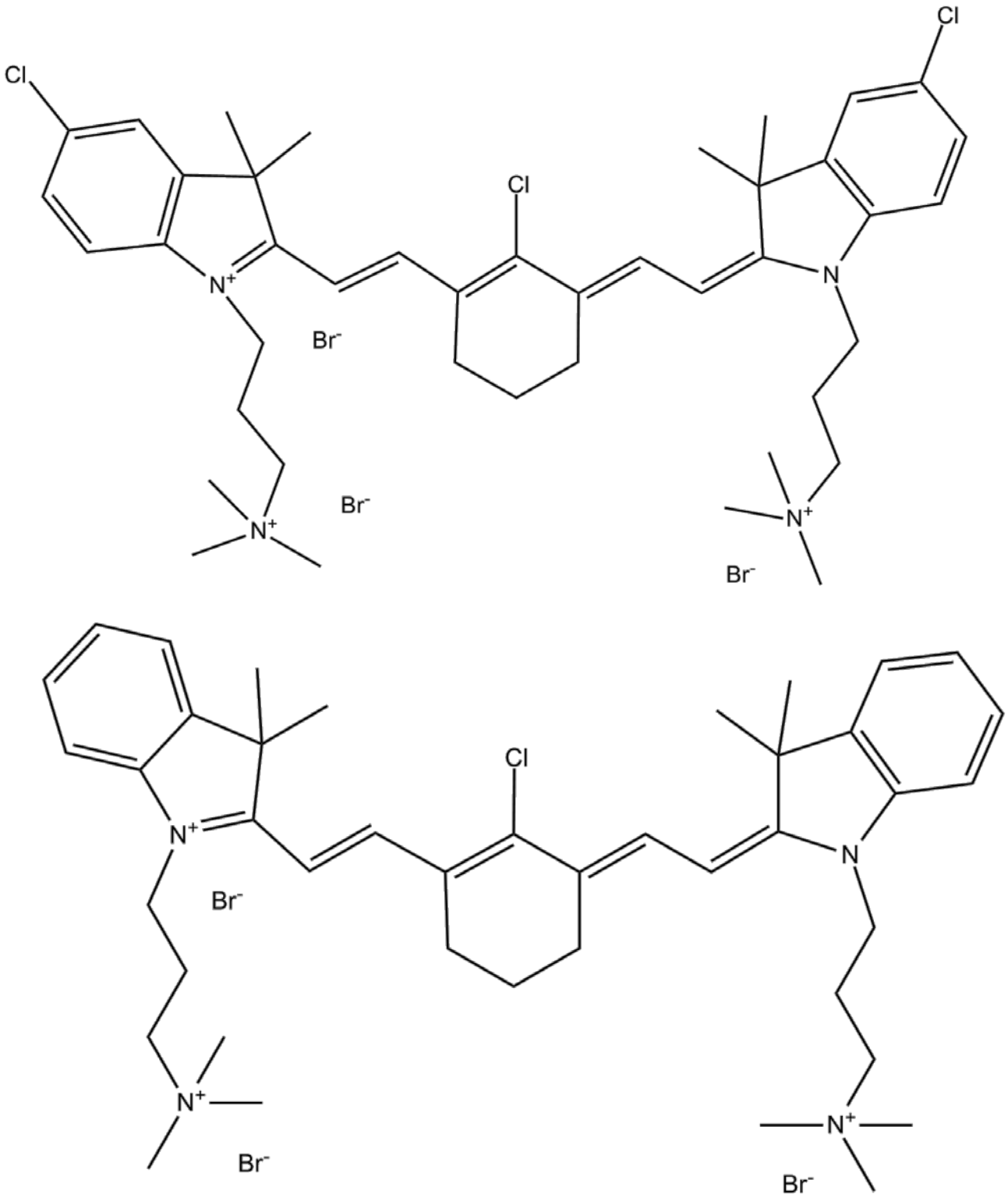


Figure 3.13: Drugs EAO-108 (top) and MHI-21 (bottom).

Both compounds are symmetric heptamethine cyanines. Both molecules exhibit a conjugated double bond character. Each compound has three positive charges. Two charges are located on the alkyl chain moieties, while the third positive charge is delocalized.

Three dimensional representations of EAO-108 and MHI-21 are illustrated in Figure 3.14 and Figure 3.15. The models were built in UCSF Chimera software package [23]. Both molecules in Figure 3.13 are represented as if they were flipped by a 90^{o} angle towards the viewer. The central chlorine groups are pointing towards the observer. The conjugated ring system makes both molecules flat. A slight concave-up curvature, clearly observed in the space-filling models, is of notice. Chlorine groups, although not very big, make the in-plane size of the EAO-108 slightly larger.

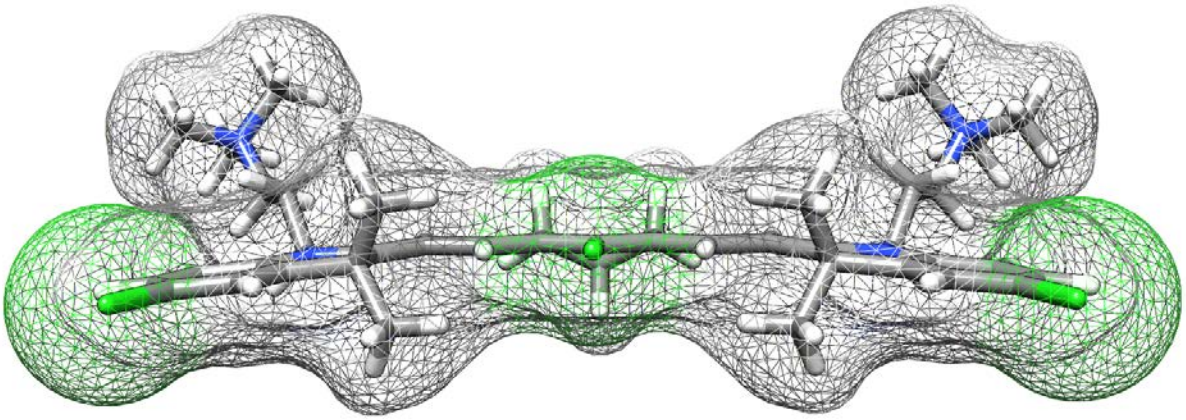


Figure 3.14: Space-filling model of EAO-108.

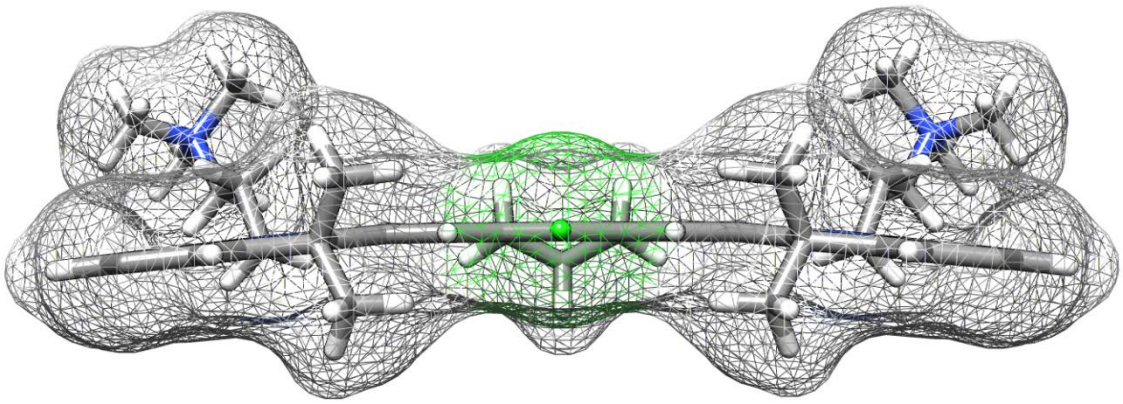


Figure 3.15: Space-filling model of MHI-21.

A complete set of data was acquired by titrating each drug individually with TEL-24, PU-22, and BCL-2-MID quadruplexes.

3.3.1.1 Titration with TEL-24

Figure 3.16 illustrates the difference in the interaction of EAO-108 and MHI-21 with the hybrid-1 of TEL-24. Surprisingly, EAO-108 readily interacts with the bottom face (3, 21, 17, and 9) of the quadruplex. To the contrary, the MHI-21 does not bind well at 1:1 mole ratio. The only imino proton signal that appears to be affected the greatest is G17, while G3 and G22 are moderately affected. Figure 3.17 illustrates additional titration experiment that was performed to confirm behavior of G17 and G4 of the TEL-24 titrated with MHI-21. At 288 K, peaks are not overlapped and the identity of each peak can be traced. At 1:0.5 mole ratio, the intensity of G17 is significantly smaller than that of G4.

When the drug is present in excess, at the 1:3 mole ratio, the mode of binding (bottom) does not change significantly. To the contrary, the MHI-21 does not bind well at 1:1 mole ratio, while more interaction with the bottom tetrad is observed at 1:3 ratio (Figure 3.18).

The color-coded models suggest that the drugs interact with the bottom tetrad and possibly with the groove of the quadruplex. Notably, that groove is the most flexible because loops neither occupy the groove (propeller), nor constrain it (lateral). However, the exact mode of interaction remains unclear.

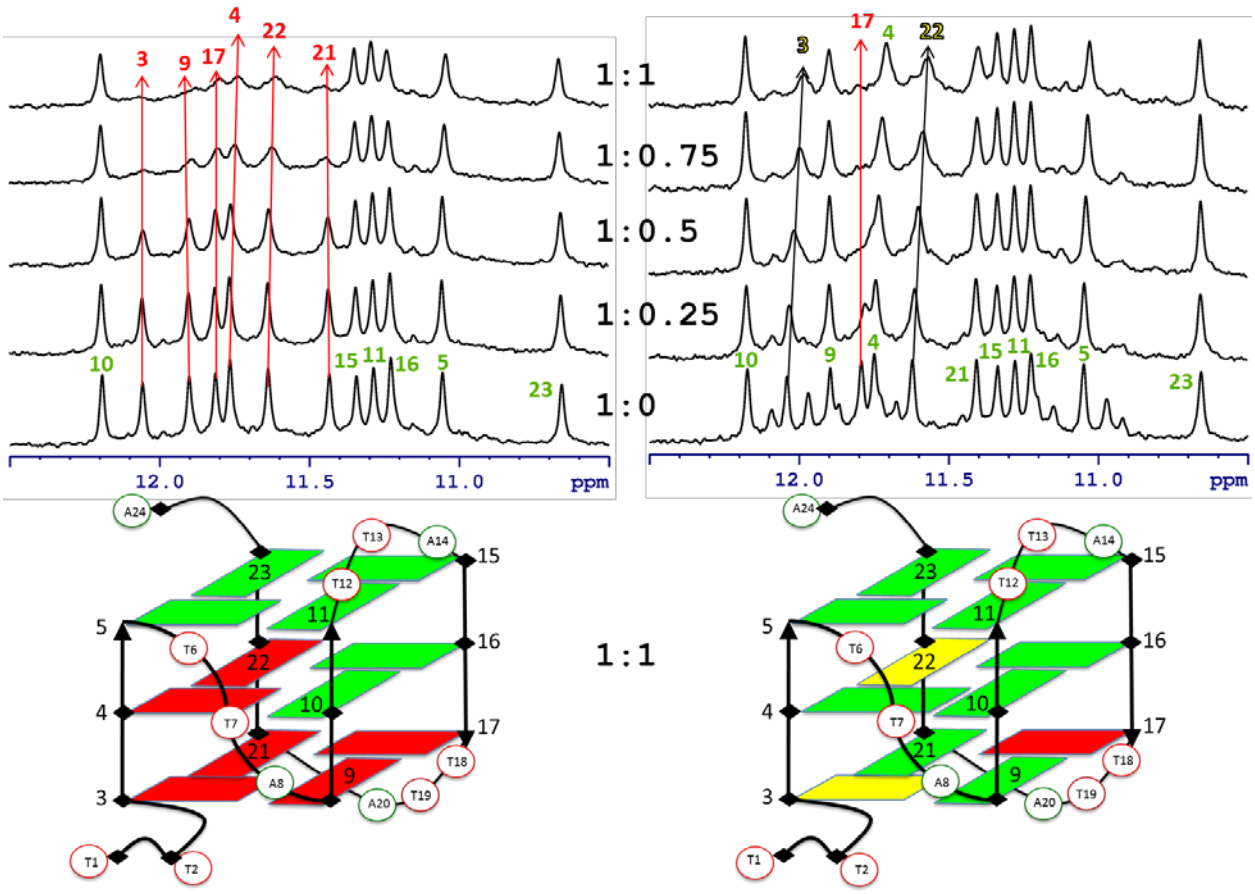


Figure 3.16: TEL-24 quadruplex titration with EAO-108 (left), and MHI-21 (right).
NMR imino proton spectra (T=298 K) of the TEL-24 quadruplex are shown on the top of the figure. EAO-108 and MHI-21 drugs were added to a constant amount of TEL-24 indicated in terms of their mole ratios. The DNA is mostly present as a hybrid-1 form in solution. Ligands were added to the quadruplex at the ratio indicated on the plot, where 1:0 ratio corresponds to the imino proton spectrum of TEL-24 with no ligand added. Schematic structures of TEL-24 are color coded according to the tetrad guanine imino protons affected in the titrations. Red residues are strongly affected, yellow are somewhat affected, and green are not affected.

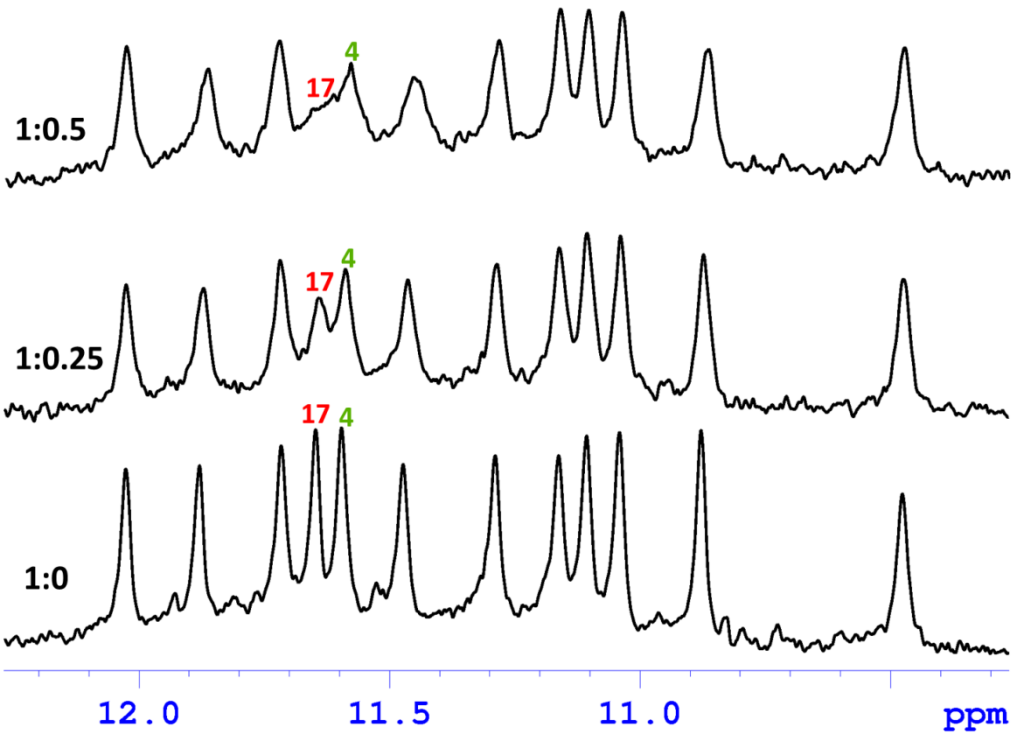


Figure 3.17: TEL-24 quadruplex titration with MHI-21 at T=288 K.

At T=288 K imino proton signals for G17 and G4 exhibit minimal overlap at 1:0. Titration to 1:0.5 mole ratio clearly shows that G17 chemical shift disappears, while the G4 imino proton signal remains.

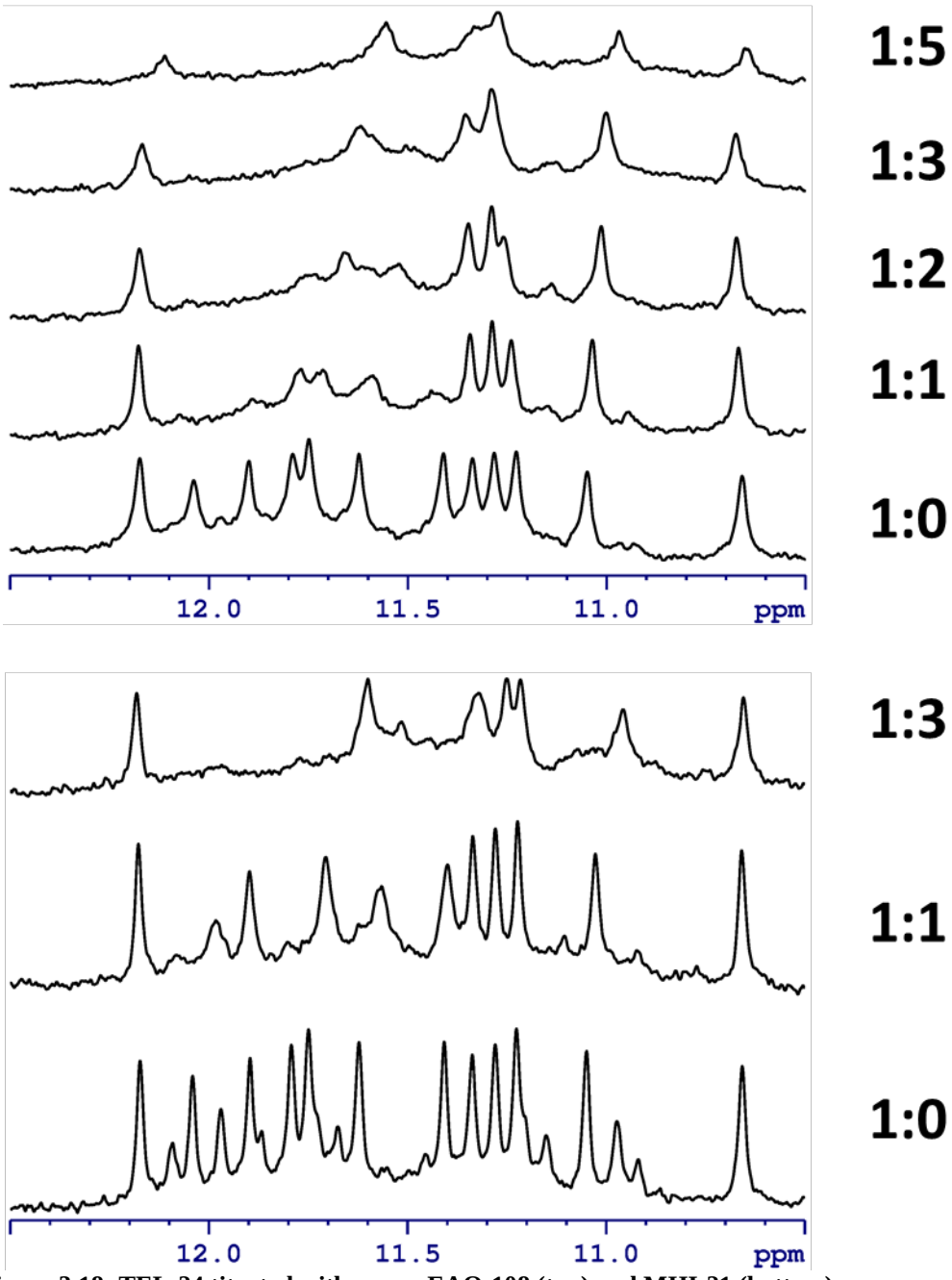


Figure 3.18: TEL-24 titrated with excess EAO-108 (top) and MHI-21 (bottom).

TEL-24 signals are broad when drugs are present in excess, which makes the analysis of the binding mechanism less precise. In general, both drugs preferentially interact with the bottom tetrad of the quadruplex.

3.3.1.2 Titration with PU-22 and BCL-2-MID

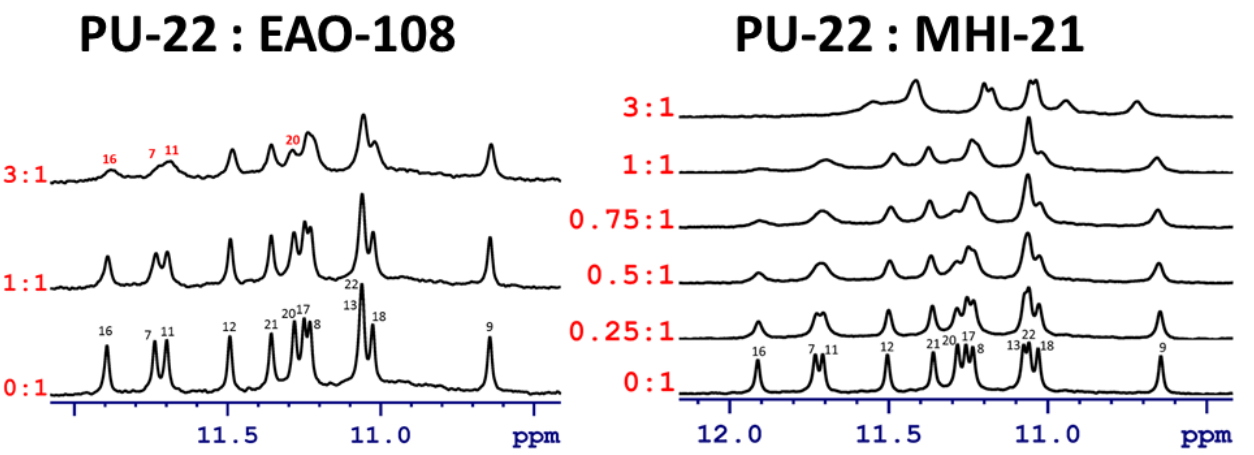


Figure 3.19: PU-22 titrated with EAO-108 and MHI-21.
PU-22 weakly interacts with both drugs. Both molecules mostly target the bottom tetrad of PU-22 (7, 11, 16, 20).

Non-telomeric G-Quadruplexes were also probed for interaction with the drugs. Titration of PU-22 shown in Figure 3.19 suggests weak interaction with the bottom tetrad of the parallel PU-22 G-quadruplex.

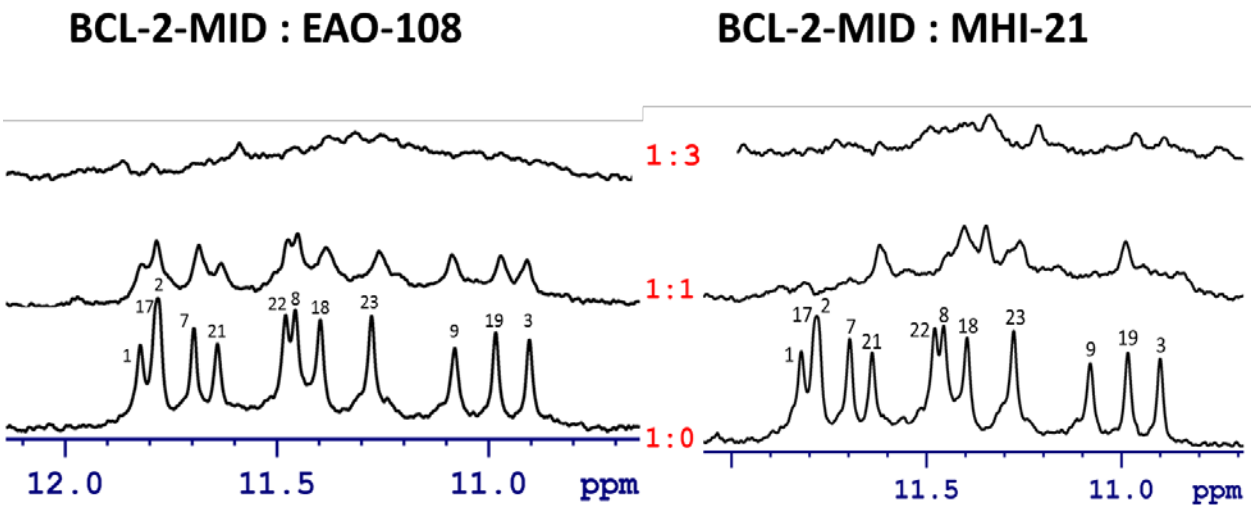


Figure 3.20: Titration of BCL-2-MID Quadruplex with EAO-108 and MHI-21.
MHI-21 exhibits some interaction with 9, 1 and 17, which belong to the bottom tetrad of the quadruplex. At 1:3 ratio signal becomes unreadable due to possible precipitation. Interaction of BCL-2-MID with EAO-108 makes all imino proton signals broader.

BCL-2-MID, an antiparallel 1xD, 2xL G-Quadruplex, is similar to TEL-24 in the strand polarity and loop arrangement. However, BCL-2-MID interacts with the EAO-108 compound differently (Figure 3.20).

These preliminary studies suggest that the EAO-108 is capable of a well-localized binding to a telomeric TEL-24 hybrid-1 G-Quadruplex, while discriminating against non-telomeric G-quadruplexes such as PU-22 and BCL-2-MID.

The ability of a small molecule to discriminate between different types of G-Quadruplexes can have a tremendous effect for potential G-Quadruplex therapeutics. For example, Dhamodharan et al. recently found that bisbenzimidazole carboxamide can discriminate c-myc-related G-Quadruplex from other telomeric G-Quadruplexes, including the parallel telomeric G-Quadruplex [24]. Although the parallel telomeric G-Quadruplex is thought to be irrelevant under physiological conditions, the discriminating ability of that small molecule is stunning.

Similarly, the properties of the EAO-108 should also be closely followed. Exact binding modes and grounds for selectivity towards the telomeric G-Quadruplex should be explored. Currently, Mass Spectrometry and Surface Plasmon Resonance experiments are underway to confirm and augment these preliminary NMR results.

As described in sections 3.1.3 and 3.1.4, an established annealing procedure has to be followed to achieve a conformationally pure sample of a G-Quadruplex. One of the pitfalls of these binding experiments is a possible incorrect folding of the G-Quadruplexes. For example, the DNA samples prepared for Mass Spectroscopy experiments have to be checked by NMR to ensure correct folding. In addition, the solubility of the drug in water and its stability in solution are crucial factors to consider when interactions with DNA are assessed. Because the DNA

imino protons are monitored in the NMR titration experiments, degradation of the small molecules can remain unnoticed and inevitably rendering the binding data invalid.

For example, Figure 3.21 depicts an NMR spectrum of EAO-108 molecules immediately after it was dissolved in 100% D_2O. The spectrum is pure, and all protons can be accounted for. However, if the solution is kept for 13 days at room temperature, it degrades by almost 50%, which is demonstrated in Figure 3.22.

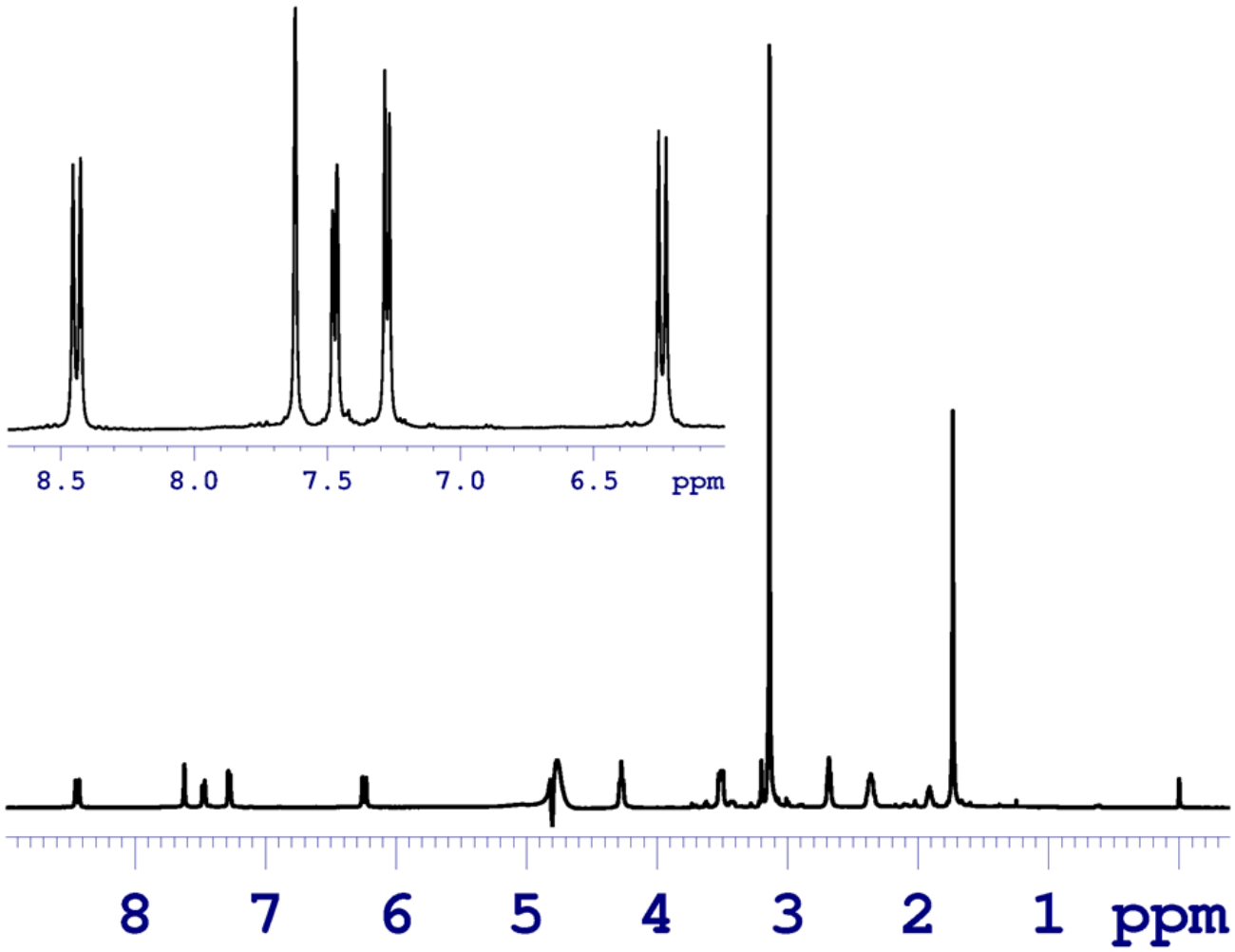


Figure 3.21: Spectrum of EAO-108 immediately after powder was dissolved in 100% D_2O.
No degradation is visible and all proton signals are accounted for.

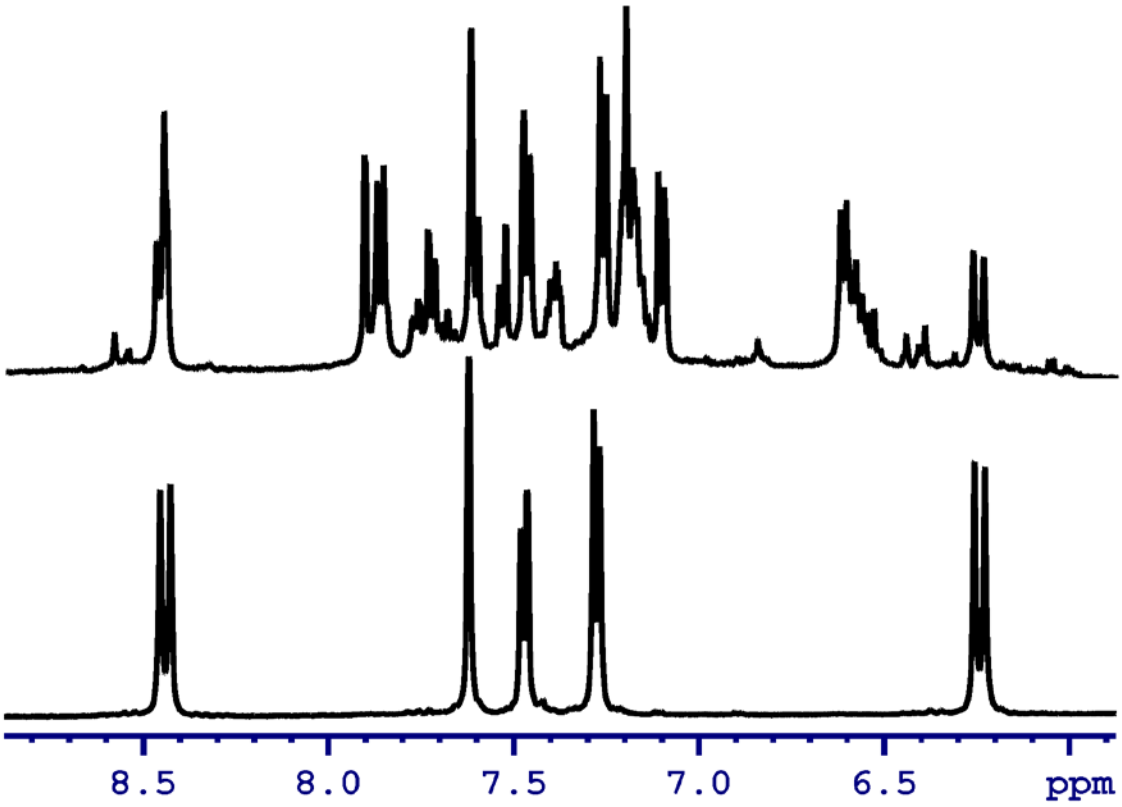


Figure 3.22: Degradation of EAO-108 monitored over two week period.
The bottom spectrum was taken immediately after the drug was dissolved in D_2O and the top spectrum was taken thirteen days later. Solution was consistently kept at room temperature. At day thirteen, at least 50% of the drug had degraded.

3.4 Targeting Different Forms of a G-Quadruplex

This section presents two compounds (ZK-26 and A-150) which were found to be selective towards the minor hybrid-2 form of hybrid-1/hybrid-2 TEL-24 mixture. Their interaction with PU-22 is also discussed.

3.4.1 ZK-26

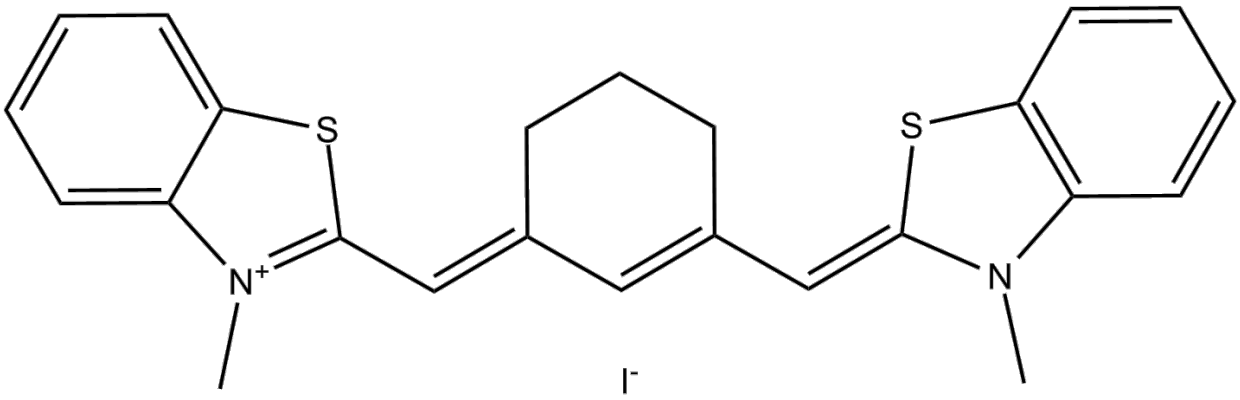


Figure 3.23: ZK-26 ontains sulfur in the five-membered rings.
The solubility in water is very poor, soluble in 100% DMSO.

ZK-26 has a very poor solubility in water, so the binding studies were performed using the 100% DMSO solution of ZK-26. Since the presence of DMSO in the DNA sample may alter the structure of DNA, the total amount of DMSO added to the sample (not more than 10% by volume) was carefully monitored. Using a separate sample of DNA, we established that the imino protons were not affected by the presence of 10% DMSO (the data is not shown here). This proves that the presence of DMSO does not alter the DNA structure.

Similar to the easily degradable EAO-108 (section 3.3.1), ZK-26 also degrades in solution. In addition, the very poor solubility of the drug required us to confirm the concentration of the ZK-26 in solution. NMR spectrum of ZK-26 was acquired, and the method of internal standard was used to confirm the actual concentration of the drug in solution (described in

Appendix A). Mole ratios in the binding studies were refined according to the actual concentration ZK-26.

3.4.1.1 Titration of ZK-26 with PU-22

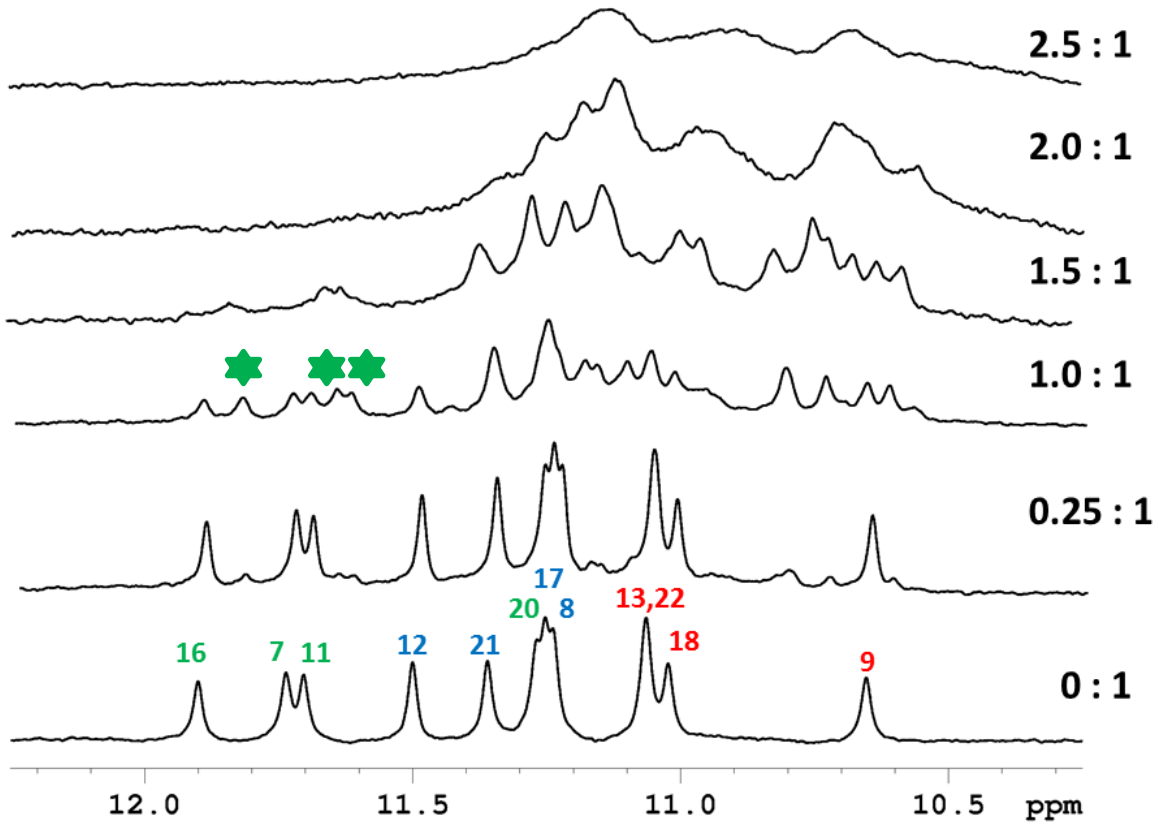


Figure 3.24: Titration of ZK-26 with PU-22.
The bottom spectrum illustrates the imino protons of the parallel form of PU-22. Peaks are labeled according to the assignment in Figure 3.1. Peaks belonging to the same tetrad are colored the same (Figure 1.3).

Surprisingly, a very strong binding is observed. Two sets of peaks are clearly visible at 1:1 mole ratio. Unlike the previous titration data (Figure 3.11, chemical shift changes; and Figure 3.16, decrease in intensities of some peaks), ZK-26-PU-22 complex stays bound long enough for NMR to detect two DNA species. The imino protons of the unbound PU-22 G-quadruplex have

the same chemical shifts as in the 0:1 spectrum. The second set of peaks, however, corresponds to the imino proton signals of PU-22 bound with ZK-26. Since the chemical environment of the imino protons is altered by the presence of the ZK-26 molecule, their chemical shift is different.

Docking studies are underway to gain additional insight in the actual mode of binding. The preliminary assessment of 1:1 spectrum suggests that the ZK-26 might be binding in the groove of PU-22 G-Quadruplex. Three grooves of the parallel PU-22 are occupied with the propeller loops. Only one groove is unoccupied and it measures 10.3 Å in size.

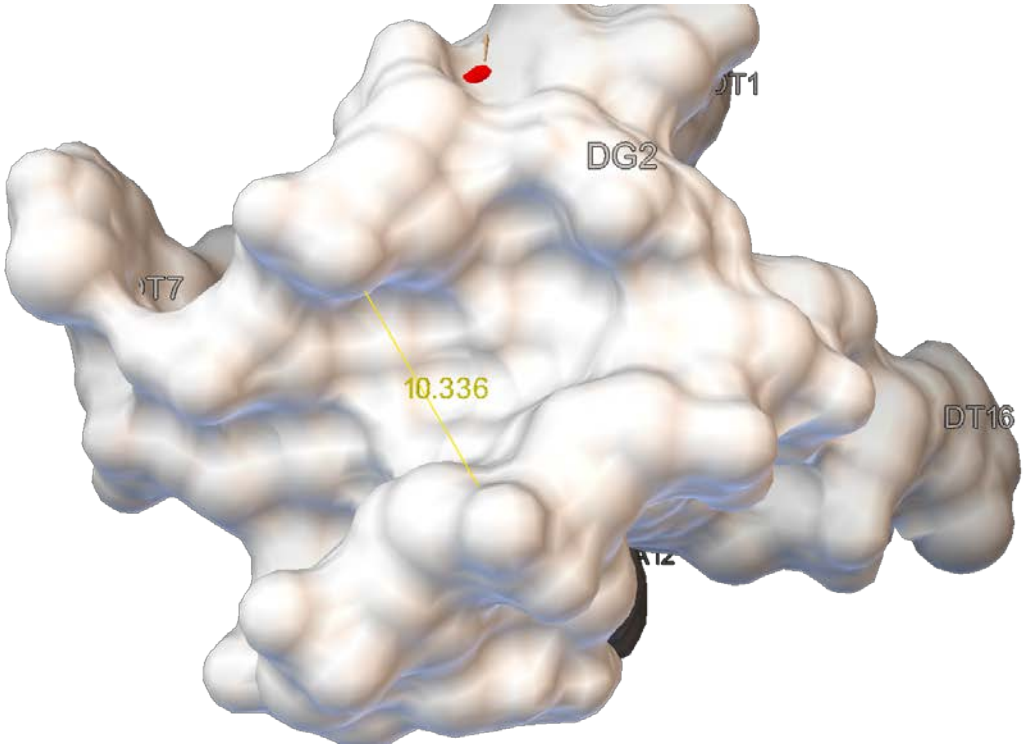


Figure 3.25: Size of the PU-22 unoccupied groove.
Some bases are labeled (DG2 – guanine base in position 2, DT16 – thymine base in position 16).

3.4.1.2 Titration of ZK-26 with TEL-24

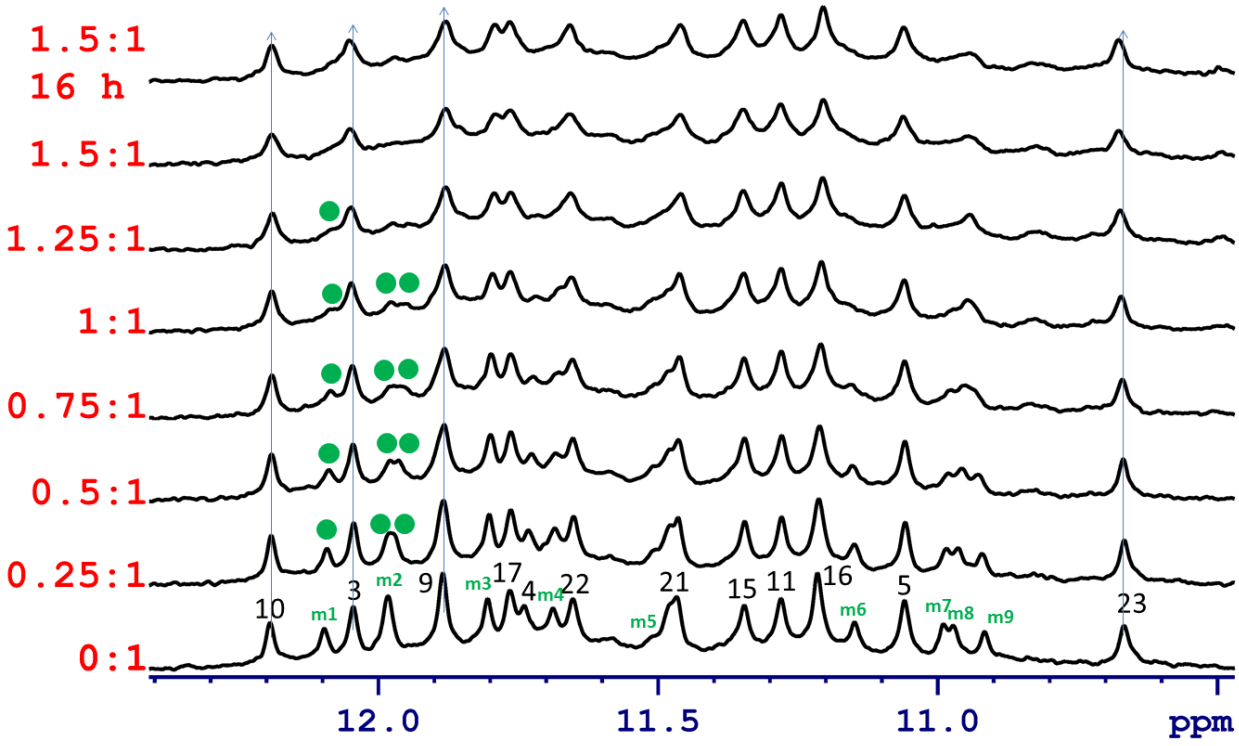


Figure 3.26: Titration of ZK-26 with TEL-24.
ZK-26 preferentially binds to the minor form of TEL-24 labeled with $m_{1—9}$. Correct assignment of the hybrid-2 TEL-24 form is now available from [12]. Changes in the chemical shift of the minor form peaks are tracked with green circles. Major form of TEL-24 is labeled in black and blue guides are shown as a vertical reference.

ZK-26 exhibits preferential binding with the hybrid-2 form of TEL-24. The titration was performed within 6 hours after the induction of the major/minor form of TEL-24. The last spectrum was acquired 22 hours after the induction.

3.4.2 Modified-ZK-26

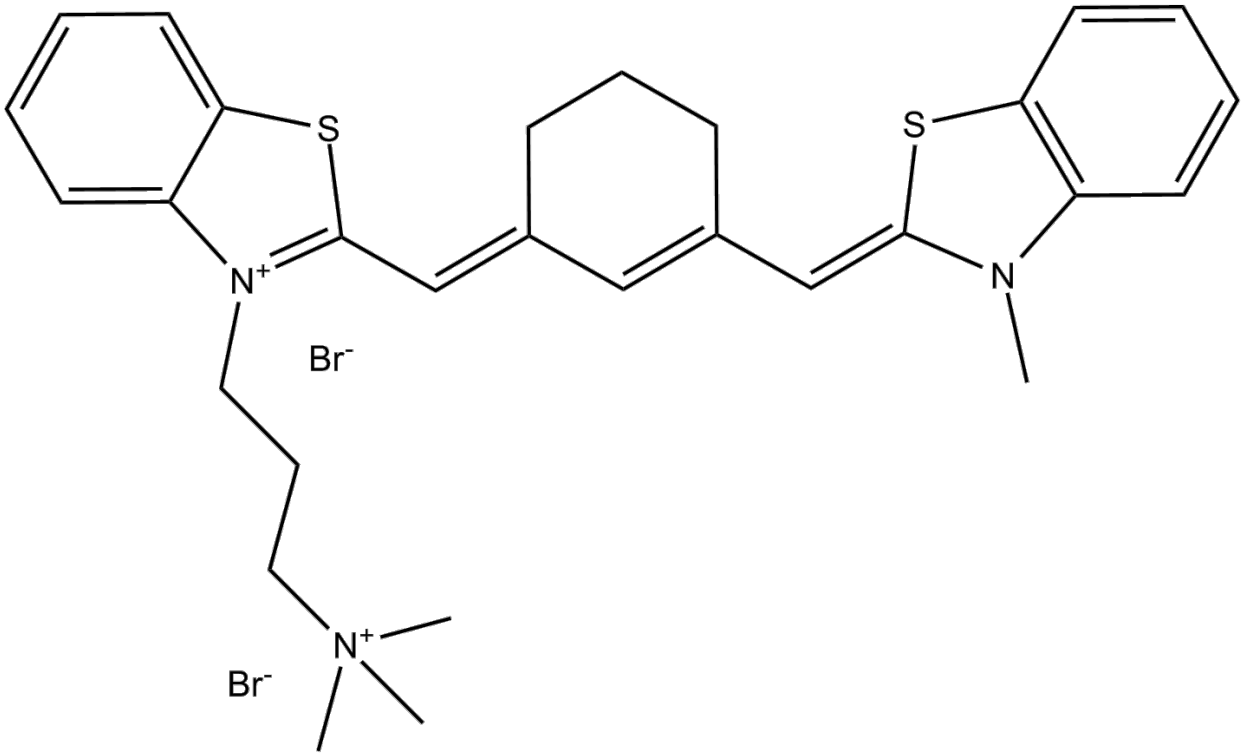


Figure 3.27: Modified ZK-26 (ZK-306).
An alkyl chain moiety was added to a molecule of ZK-26 in order to increase its water solubility. However, only slightly enhanced water solubility was observed. Binding studies were performed using a DMSO solution of the modified-ZK-26.

Even though the alkyl modification slightly improved water solubility, the complex aggregates at a 3:1 mole ratio. The binding of the modified-ZK-26 with the PU-22 is still strong, which is evident from the two sets of peaks at 1:1.

3.4.2.1 Titration of Modified-ZK-26 with PU-22

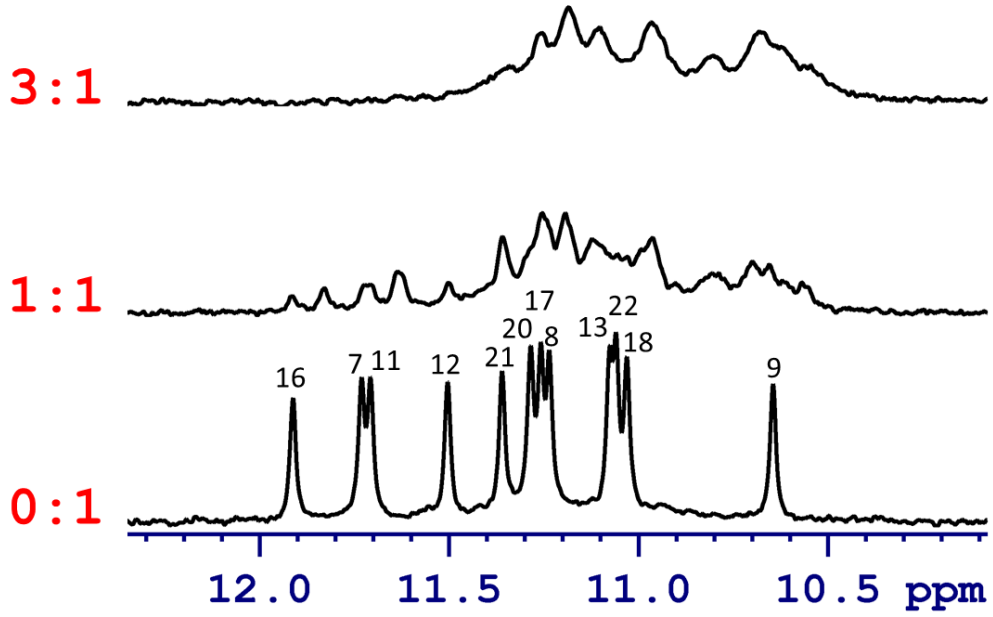


Figure 3.28: Titration of PU-22 with Modified-ZK-26.

Although the NMR spectrum quality is poor, the result is intriguing. Technically, aggregation of the quadruplexes might also support inhibition of the telomerase and may be a mechanism of the successful therapeutics agent.

3.4.2.2 Titration of Modified-ZK-26 with TEL-24

Modified-ZK-26 also binds to the minor, hybrid-2, form of the TEL-24. The major form is clearly left intact.

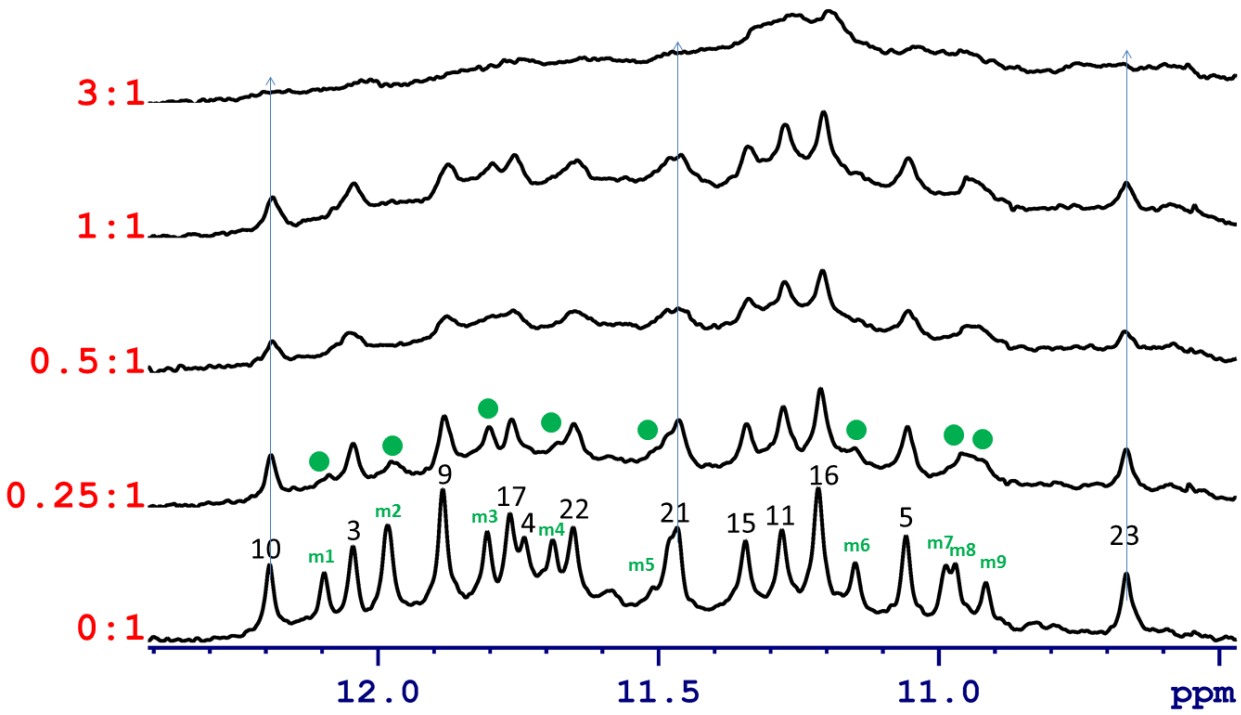


Figure 3.29: Titration of TEL-24 with Modified-ZK-26.
Minor form of TEL-24 was induced according to the standard method. Major form is not affected by the binding, while hybrid-2 form labeled with $m_{1\text{-}9}$ in green is affected. However, already at 0.5:1 ration the complex aggregates and starts to precipitate out of solution.

The binding abilities of the modified-ZK-26 are similar to that of the ZK-26. Alkyl moiety of the modified-ZK-26 increased aggregation rates of the DNA-drug complexes. Although not suitable for the NMR structure work, the binding mode might present useful possibilities for therapeutics development.

3.4.3 A-150

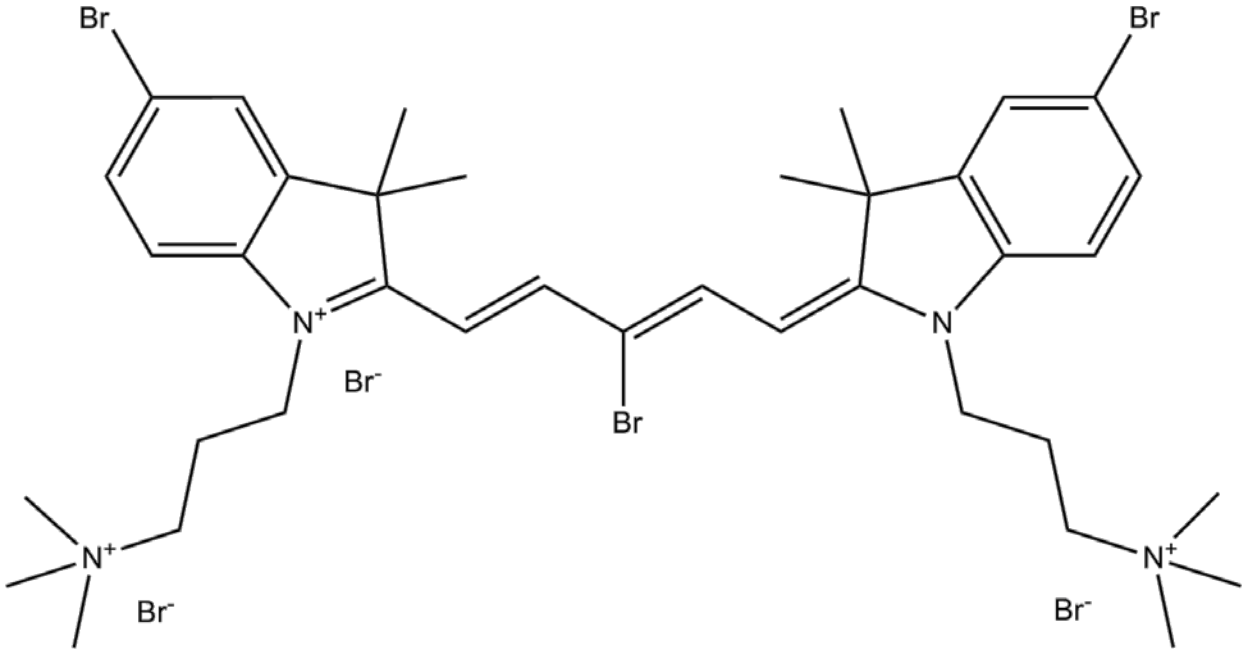


Figure 3.30: Structure of A-150 drug.
Two bulky bromine groups are attached to the six-membered rings.

A-150 is very well-soluble in water. The drug weakly interacts with the TEL-24 50/50 mixture. A-150 preferentially interacts with the hybrid-2 until 1.5:1 ratio is reached.

3.4.3.1 A-150 Titration with TEL-24

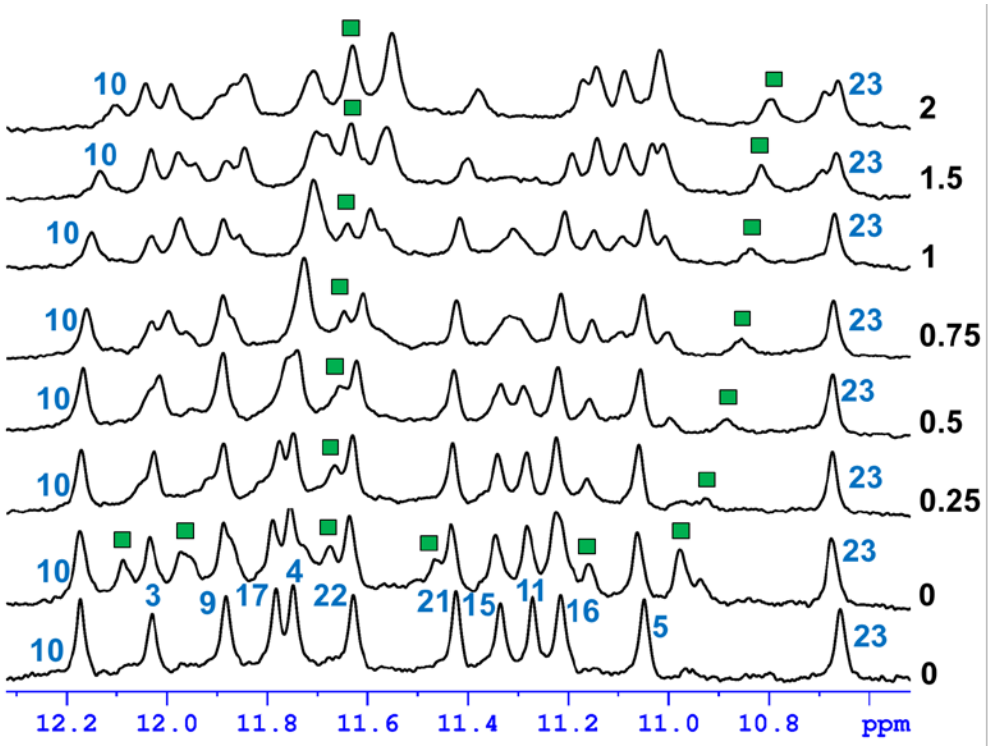


Figure 3.31: Titration of A-150 with TEL-24.
Both hybrid-1 and hybrid-2 forms of TEL-24 interact with the drug. Hybrid-2 imino proton signals, which are affected the most, are labeled with green squares.

3.4.3.2 A-150 Titration with PU-22

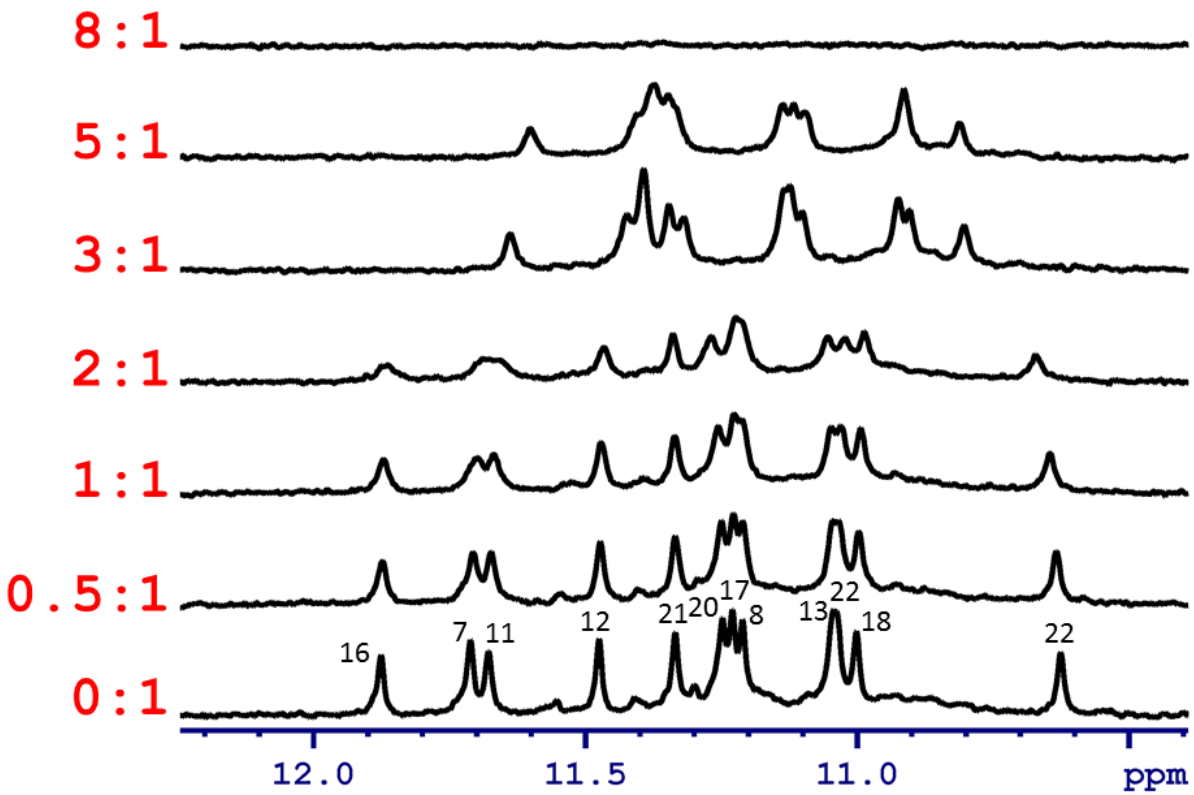


Figure 3.32: Titration of A-150 drug with PU-22 quadruplex.
Most likely the binding occurs on the top and bottom tetrad. At 8:1 ratio, the complex precipitates out of the solution.

The drug stacks to the top and the bottom tetrad of the PU-22 G-quadruplex. The complex precipitates at 8:1 mole ratio. Chemical shifts of 16 and 22 imino protons drastically change at 3:1 ratio. Titration has to be repeated with smaller increments to understand specific modes of interaction.

A-150 also interacts with BCL-2-MID, but the complex precipitates at 2:1 mole ratio.

3.4.3.3 A-150 Titration with BCL-2-MID

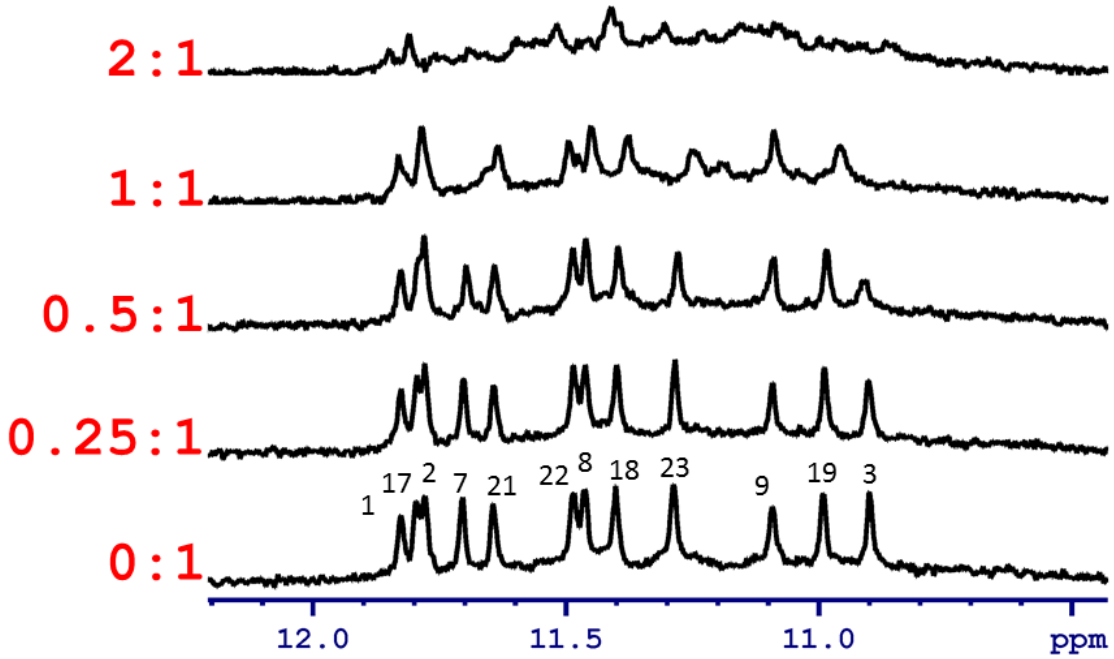


Figure 3.33: Titration of A-150 with BCL-2-MID.
At 2:1 ratio, the complex begins to precipitate.

In addition to a well-resolved imino proton spectrum of the PU-22 G-Quadruplex, its phosphorus spectrum can be used to monitor binding. Because the spectrum accounts for every phosphodiester in the molecule, it can be particularly useful to detect loop and groove interactions. Figure 3.34 displays two phosphorus spectra of PU-22. The bottom spectrum is of the PU-22 phosphate backbone. The top spectrum corresponds to the phosphate backbone of the PU-22 parallel G-quadruplex bound with the A-150 at 1:1 mole equivalents. Remarkably, all phosphate shifts can be monitored and the 1:1 spectrum is also very well-resolved.

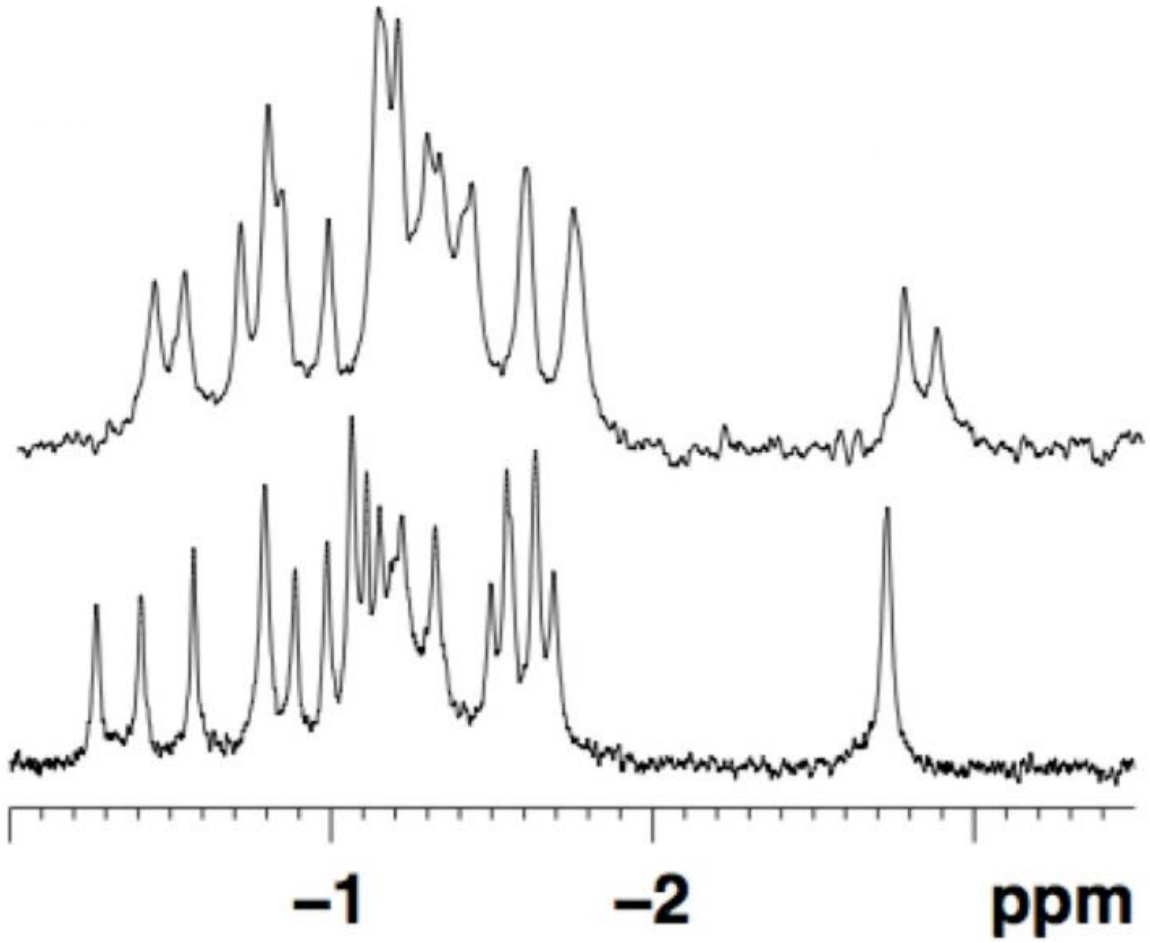


Figure 3.34: ^{31}P spectrum of PU-22 titrated with A-150 (1:1 mole ratio).
The top spectrum corresponds to PU-22 interacting with A-150.

In addition to the imino proton and the phosphate spectrum of the G-quadruplex, aromatics and sugar region of the proton spectrum contain important information about the structural perturbations and binding interactions. To simplify aromatics region of the proton spectrum, a novel proton-exchange method was developed in our laboratory and is currently under refinement.

3.5 Novel Proton Exchange Method

Guanine and adenine protons located in the eighth position (AH8 and GH8 are color coded in Figure 1.9) resonate in the 7-8 ppm region of the NMR proton spectrum. In addition to AH8 and GH8 protons, AH2, C and T H6 are also found in that region. Similarity in the chemical shifts of these protons leads to significant signal overlap. Since the structural research depends on the differentiation of the observed peaks, expensive isotope labeling is used. However, it is a costly and demanding process, which may be prohibitive for screening.

A new procedure was designed to provide an alternative simple and cost-effective method described in section 3.5.1. It was used to exchange AH8 and GH8 with deuterium, which can simplify the base region of the DNA NMR spectrum. It is especially important for G-Quadruplex DNA reach in guanine residues. Their spectra can be considerably simplified, which allows for a more accurate examination of grooves, loops, and the flanking sequences.

3.5.1 Schematic representation of the method

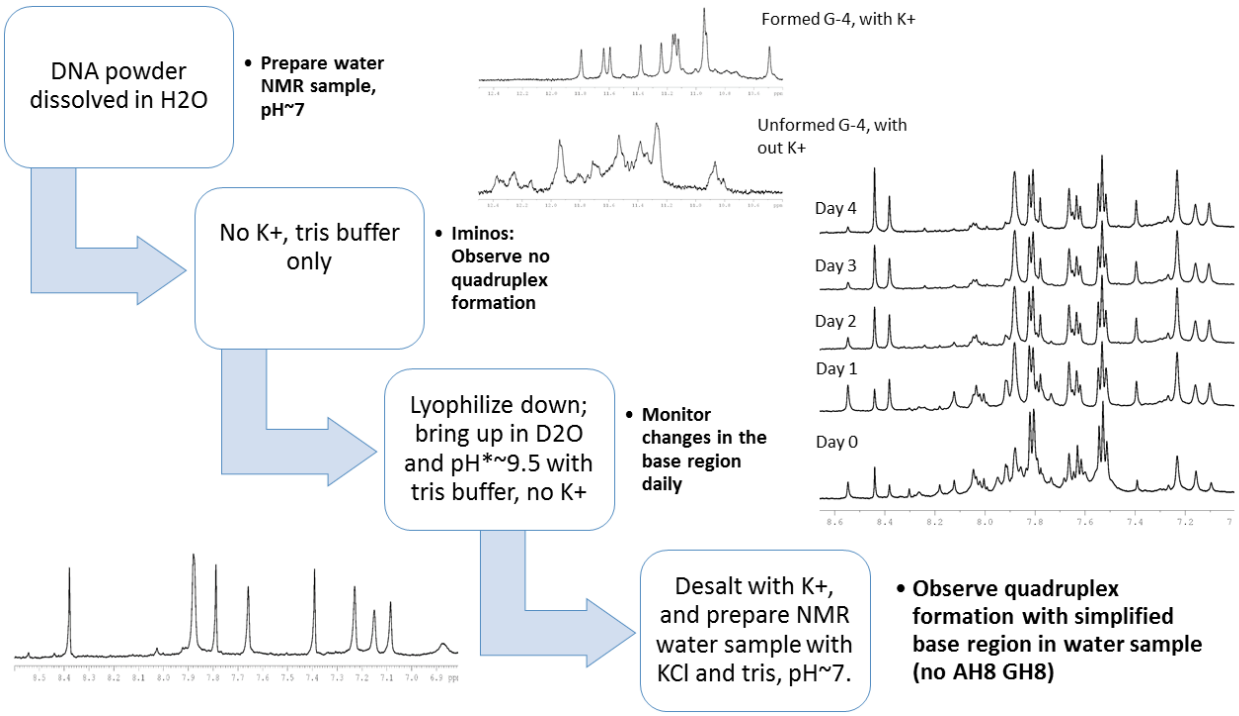


Figure 3.35: Schematic representation of the novel proton exchange method.
Spectra shown are based on the preliminary results obtained from a PU-22 proton-exchanged sample.

The DNA sample is prepared in a solution of only TRIS-HCl buffer and water. The reference spectrum is then recorded to observe the lack of organized structure. The solvent is then exchanged with D_2O and pH* is raised to approximately 9.5. It is important to note that the pH is raised using the TRIS-base in D_2O. A commonly used KOH and NaOH bases should be avoided, because introduction of potassium and sodium cations might trigger and stabilize the G-Quadruplex folding. The sample is then kept at an elevated temperature (65-55 ^oC) to speed up the exchange. Extent and completion of exchange can be monitored over time. A fully exchanged sample is then desalted with the desired cation and pH is adjusted back to 7. Non-destructiveness of the method can be easily proved by recording an imino proton spectrum in

water or observing phosphodiester backbone. A selectively labeled DNA sample can then be used for structural and binding studies.

3.5.2 H+ exchange of Dickerson decamer

A Dickerson decamer, a B-DNA self-complimentary 10-mer sequence, $d(GCGAATTCGC)_2$, was used as a positive control. Imino and base regions of the proton NMR spectrum of this well-studied 10-mer DNA were previously assigned and are shown in the bottom spectrum in Figure 3.36 [25]. The DNA sample AH8 and GH8 protons were selectively labeled with deuterium, which is evident from the top spectrum in the figure.

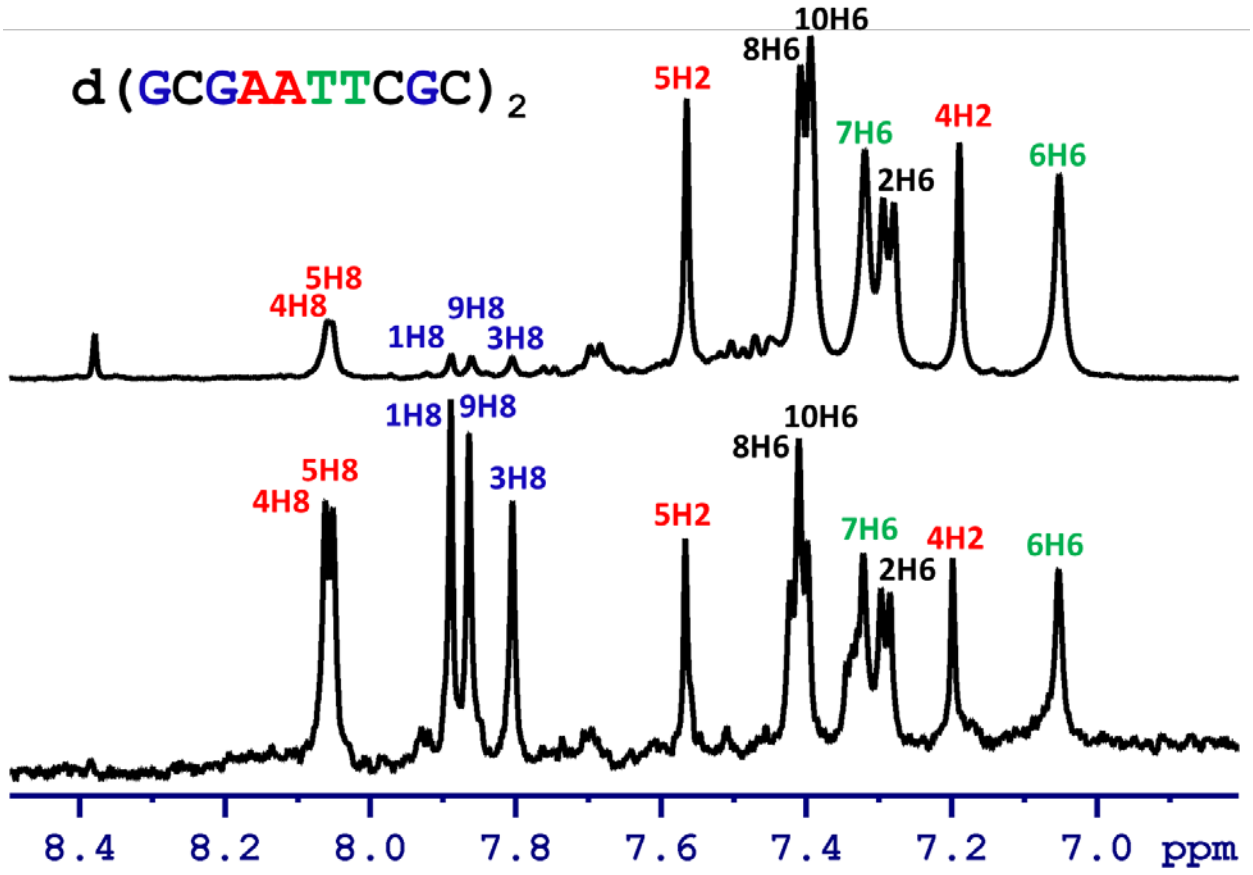


Figure 3.36: Proton exchange of the Dickerson-Decamer DNA.
The bottom spectrum shows a spectrum of the control 10-mer with the protons labeled according to the assignment published by [25]. The top spectrum is of a proton-exchange 10-mer sample, where AH8 and GH8 signals are significantly reduced.

Preliminary results were also obtained for PU-22 G-Quadruplex. Figure 3.37 compares a base region of the proton spectrum of control PU-22 to that of the labeled sample. When all GH8 and AH8 signals are silenced, only AH2 and TH6 proton resonances are left. Structural perturbations of the loops and grooves (rich in A and T residues) can then be readily monitored. Currently, we are in the process of labeling other G-Quadruplex DNA samples.

3.5.3 H+ exchange of PU-22

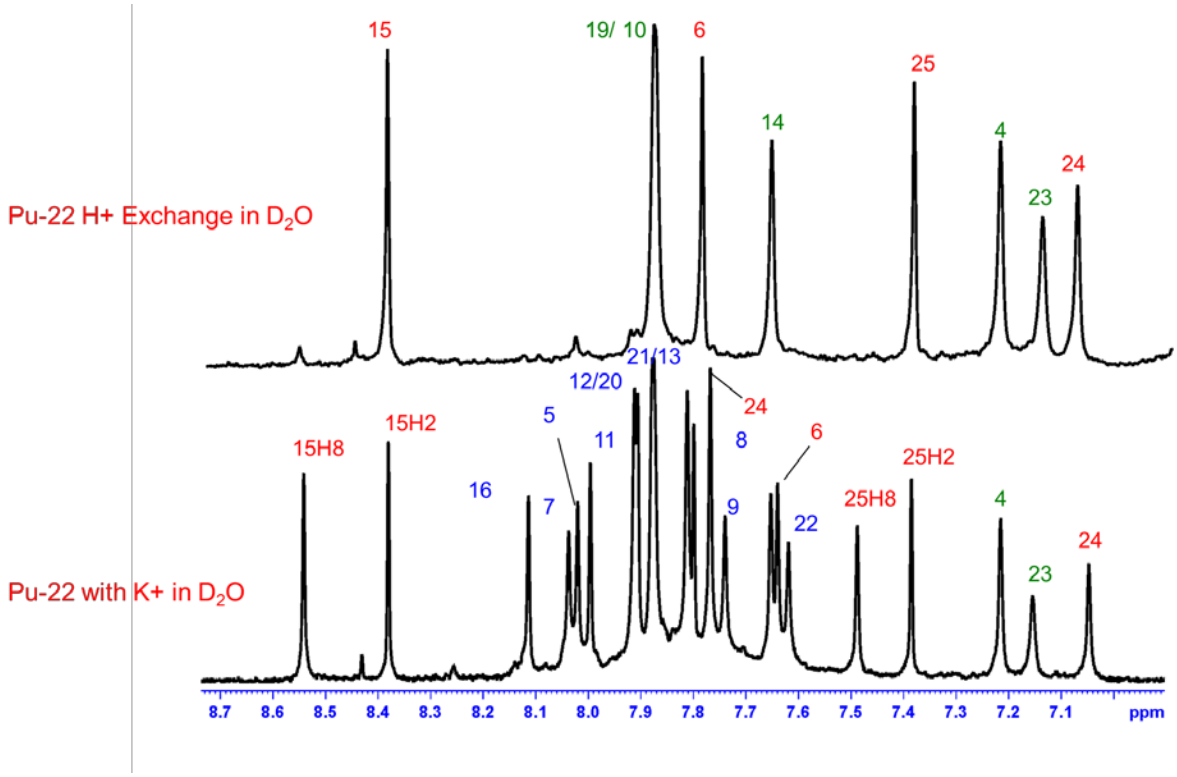


Figure 3.37: Preliminary studies on the PU-22 G-Quadruplex labeling.
Bottom spectrum is of the aromatic region of PU-22 folded into the parallel G-Quadruplex ([7]). Guanine H8 are labeled in blue, adenine H8 and H2 are labeled in red, and thymine H6 are labeled in green. Top spectrum is the same PU-22 folded (we proved that showing the imino protons of the exchange sample briefly put in water) G-Quadruplex. Only remaining peaks are labeled (red: AH2 and green: TH6).

4 CONCLUSIONS

G-Quadruplex DNA is implicated in a variety of cancers and significant insights into the structure and function of G-Quadruplexes have been made over the years. Even though

development of therapeutics which could specifically target G-Quadruplexes is underway, no clinically approved drugs are available to date. In this work, differences in structural and folding patterns of representative G-Quadruplexes were studied, and new methods were developed to make research on G-Quadruplexes less complex and affordable. A library of potential drug candidates was screened. We have identified small molecules that have the ability to discriminate between topologies and sequences of the G-Quadruplex DNA, were identified and tested. Even though the preliminary results were very promising, more research is needed to understand the underlying mechanisms of interaction. For example, binding sites can be localized more precisely with NOE contacts to specific G-Quadruplex residues. Also, chemical shift perturbations of GH8 and H1' protons can be monitored. Because those protons point into the grooves of the G-Quadruplexes, they serve as a convenient reporter for groove interactions. In addition, a novel labeling procedure developed here will also facilitate structural and drug interaction studies due to significantly minimized NMR shift overlaps. Although GH8 and AH8 signals will be silenced in this method, loop TH6 resonances can be conveniently utilized instead. Binding sites could also be further studied with ^{19}F NMR. Labeled nucleotides can be synthesized in the lab at a lowered cost. Thymidine residues present in abundance in the G-Quadruplex loops can be substituted with ^{19}F labeled deoxy-uridine. Substitution of thymine present in the capping loops will allow for monitoring of the stacking interactions, while labeled core loops will monitor the loop binding. In addition, more studies are necessary to assess the biological efficacy and the mechanism of action of these drug candidates in a cell environment.

REFERENCES

[1] D. Davies, G. Felsenfeld, The structure of RNA, Structural chemistry and molecular biology. WH Gree-man \& Co., San Francisco (1968) 422--429.

[2] M. Gellert, M.N. Lipsett, D.R. Davies, Helix formation by guanylic acid, Proceedings of the National Academy of Sciences of the United States of America 48 (1962) 2013-2018.

[3] E.H. Blackburn, Structure and function of telomeres, Nature 350 (1991) 569-573.

[4] S. Neidle, Therapeutic Applications of Quadruplex Nucleic Acids, Therapeutic Applications of Quadruplex Nucleic Acids (2012) 1-196.

[5] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure, Nucleic acids research 34 (2006) 5402-5415.

[6] H. Yaku, T. Murashima, D. Miyoshi, N. Sugimoto, Specific Binding of Anionic Porphyrin and Phthalocyanine to the G-Quadruplex with a Variety of in Vitro and in Vivo Applications, Molecules 17 (2012) 10586-10613.

[7] A. Ambrus, D. Chen, J. Dai, R.A. Jones, D. Yang, Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization, Biochemistry 44 (2005) 2048-2058.

[8] B.G. Kim, H.M. Evans, D.N. Dubins, T.V. Chalikian, Effects of Salt on the Stability of a G-Quadruplex from the Human c-MYC Promoter, Biochemistry 54 (2015) 3420-3430.

[9] J.X. Dai, D. Chen, R.A. Jones, L.H. Hurley, D.Z. Yang, NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region, Nucleic acids research 34 (2006) 5133-5144.

[10] H.X. Sun, J.F. Xiang, Y.H. Shi, Q.F. Yang, A.J. Guan, Q. Li, L.J. Yu, Q. Shang, H. Zhang, Y.L. Tang, G.Z. Xu, A newly identified G-quadruplex as a potential target regulating Bcl-2 expression, Biochimica Et Biophysica Acta-General Subjects 1840 (2014) 3052-3057.

[11] K.N. Luu, A.T. Phan, V. Kuryavyi, L. Lacroix, D.J. Patel, Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold, Journal of the American Chemical Society 128 (2006) 9963-9970.

[12] I. Bessi, H.R. Jonker, C. Richter, H. Schwalbe, Involvement of Long-Lived Intermediate States in the Complex Folding Pathway of the Human Telomeric G-Quadruplex, Angewandte Chemie (2015).

[13] J. Abraham Punnoose, Y. Cui, D. Koirala, P.M. Yangyuoru, C. Ghimire, P. Shrestha, H. Mao, Interaction of G-quadruplexes in the full-length 3' human telomeric overhang, Journal of the American Chemical Society 136 (2014) 18062-18069.

[14] B.J. Chen, Y.L. Wu, Y. Tanaka, W. Zhang, Small molecules targeting c-Myc oncogene: promising anti-cancer therapeutics, International journal of biological sciences 10 (2014) 1084-1096.

[15] J. Dai, T.S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R.A. Jones, D. Yang, An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution, Journal of the American Chemical Society 128 (2006) 1096-1098.

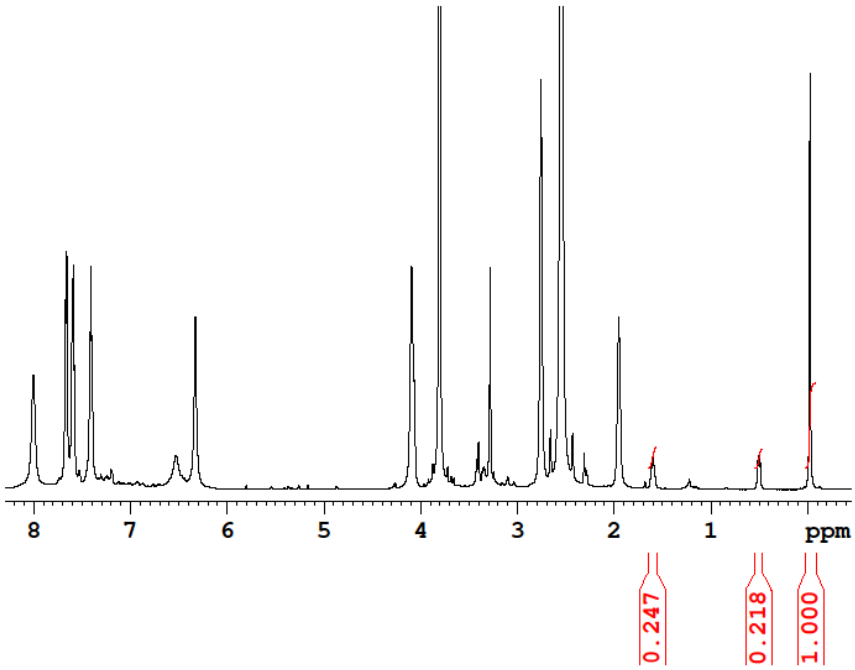
[16] V.A. Bloomfield, D.M. Crothers, I. Tinoco, Nucleic Acids: Structures, Properties, and Functions, University Science Books, 2000.

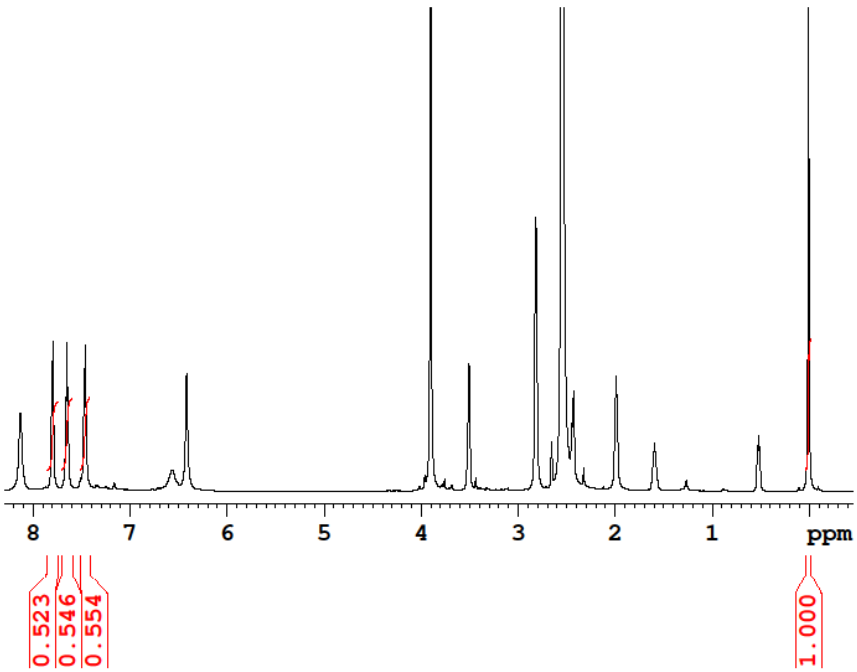
[17] R.C. Gamble, J.P. Schoemaker, Rate of tritium labeling of specific purines in relation to nucleic acid and particularly transfer RNA conformation, Biochemistry 15 (1976) 2791-2799.
[18] P. Plateau, M. Gueron, Exchangeable proton NMR without base-line distorsion, using new strong-pulse sequences, Journal of the American Chemical Society 104 (1982) 7310-7311.
[19] M.L. Liu, X.A. Mao, C.H. Ye, H. Huang, J.K. Nicholson, J.C. Lindon, Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy, Journal of Magnetic Resonance 132 (1998) 125-129.
[20] R.D. Gray, J.O. Trent, J.B. Chaires, Folding and unfolding pathways of the human telomeric G-quadruplex, Journal of molecular biology 426 (2014) 1629-1650.
[21] I.M. Kuznetsova, K.K. Turoverov, V.N. Uversky, What Macromolecular Crowding Can Do to a Protein, International Journal of Molecular Sciences 15 (2014) 23090-23140.
[22] R. Nanjunda, E.A. Owens, L. Mickelson, T.L. Dost, E.M. Stroeva, H.T. Huynh, M.W. Germann, M.M. Henary, W.D. Wilson, Selective G-quadruplex DNA recognition by a new class of designed cyanines, Molecules 18 (2013) 13588-13607.
[23] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis, Journal of computational chemistry 25 (2004) 1605-1612.
[24] V. Dhamodharan, S. Harikrishna, A.C. Bhasikuttan, P.I. Pradeepkumar, Topology specific stabilization of promoter over telomeric G-quadruplex DNAs by bisbenzimidazole carboxamide derivatives, ACS chemical biology 10 (2015) 821-833.
[25] J.M. Aramini, B.W. Kalisch, R.T. Pon, J.H. vandeSande, M.W. Germann, Structure of a DNA duplex that contains alpha-anomeric nucleotides and 3'-3' and 5'-5' phosphodiester linkages: Coexistence of parallel and antiparallel DNA, Biochemistry 35 (1996) 9355-9365.

APPENDICES

Appendix A (*Concentration Determination of Drugs ZK-26)*

1. Prepare stock solution of a drug in the desired solvent, high concentration are preferred;
2. Record approximate concentration of the drug in the stock solution;
3. Using standard solution of DSS (known concentration) and assuming approximate concentration of the drug from step 2 prepare NMR sample in D_2O.
4. Record NMR spectrum using zgpr pulse sequence;
5. Identify DSS peaks; identify major peak around 0 ppm and set it to 0.0 ppm. This is the reference peak. Also identify minor DSS peaks: triplet at 3.1 ppm, pentet at 1.9 ppm, triplet at 0.8 ppm. Intensities of each of the minor peak should be 22% of the intensity of the reference peak.
6. Integrate reference peak and set integral value to 1 for a multiplicity of 9H. Integrate two minor DSS peaks and compare to the theoretical 0.22 value for 2H. Continue if consistent.
7. Identify three peaks belonging to the drug. Identify the number of protons each peak is resonating from. Integrate each peak separately.
8. Set the integral of the reference DSS peak to 9, (reference DSS peak is coming from 9 magnetically identical protons).
9. Calculate concentration of the drug in the NMR sample and the concentration of the stock solution of the drug (assuming no pipetting error was made).
10. Example data and calculations are shown in the figures below.


8
7
6
5
4
3
2
1
ppm
0.247
0.218
1.000


8
7
6
5
4
3
2
1
ppm
0.523
0.546
0.554
1.000

Appendix B (Other G-Quadruplex drugs)

Appendix B.1 (E-46)

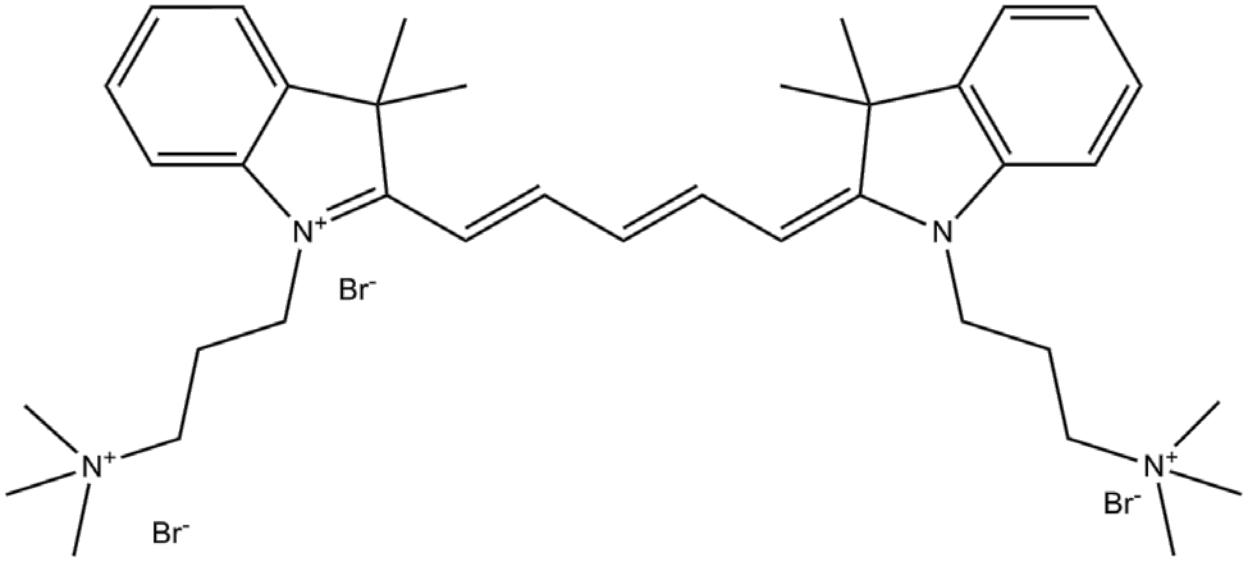


Figure 1: E-46 Cyanine Drug.
There are no bulky substituents on the six-membered rings. The linker consists of single bonds and it is flexible.

4.1.1.1 E-46 Titration with PU-22

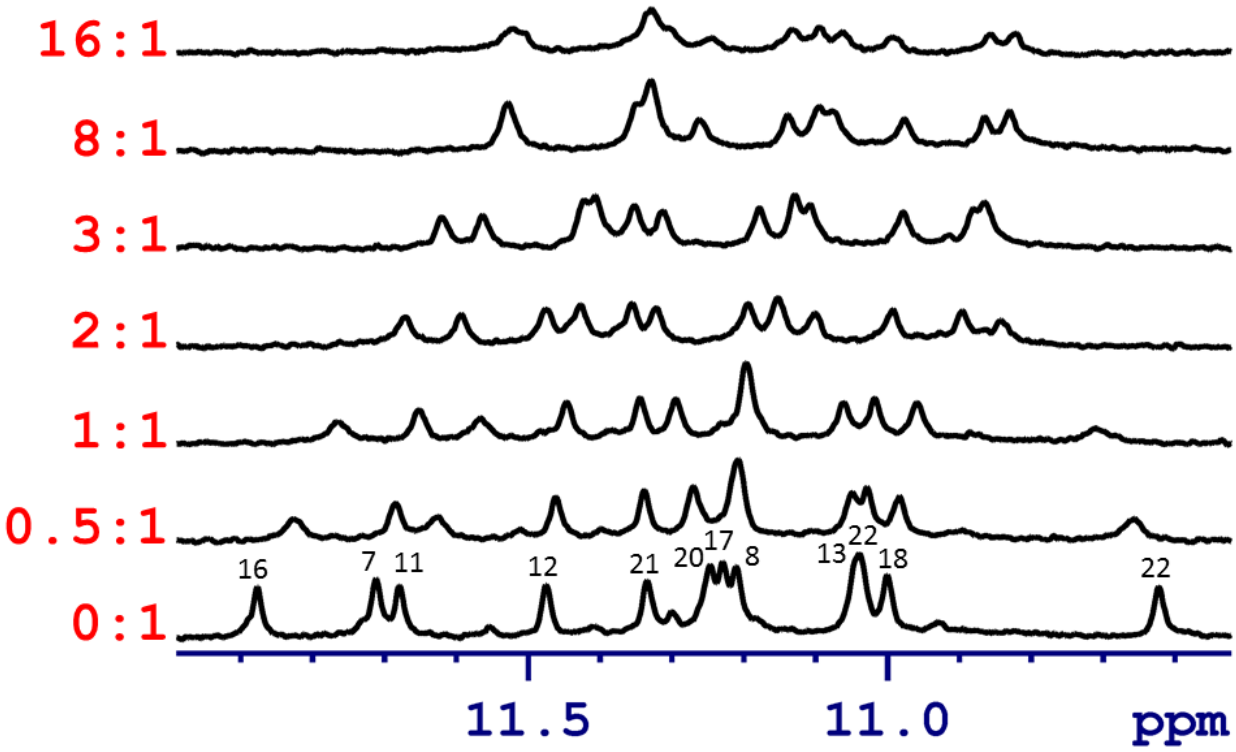


Figure 2: Titration of E-46 with PU-22.
Even at 16:1 ratio, the complex has not yet precipitated out of the solution. Binding occurs on the top and on the bottom tetrad.

Appendix B.2 (T-5)

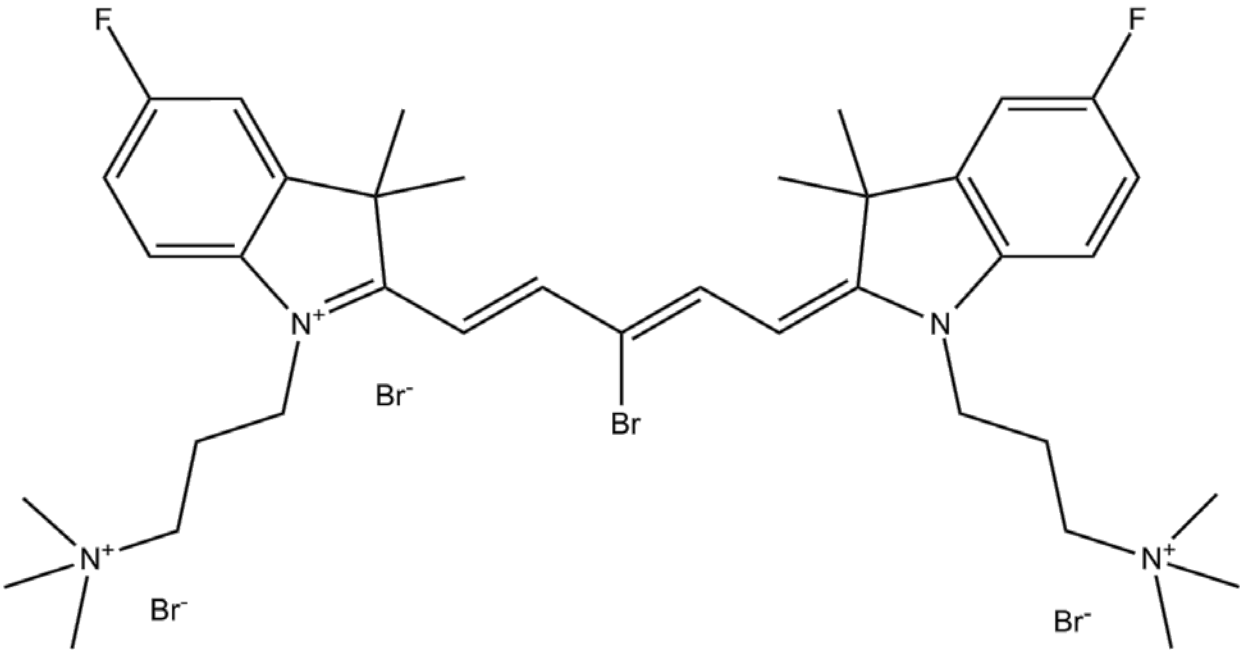


Figure 3 T-5 Cyanine Drug.
There are no bulky substituents on the six-membered rings. The linker consists of double bonds and it is not flexible.

4.1.1.2 T-5 Titration with PU-22

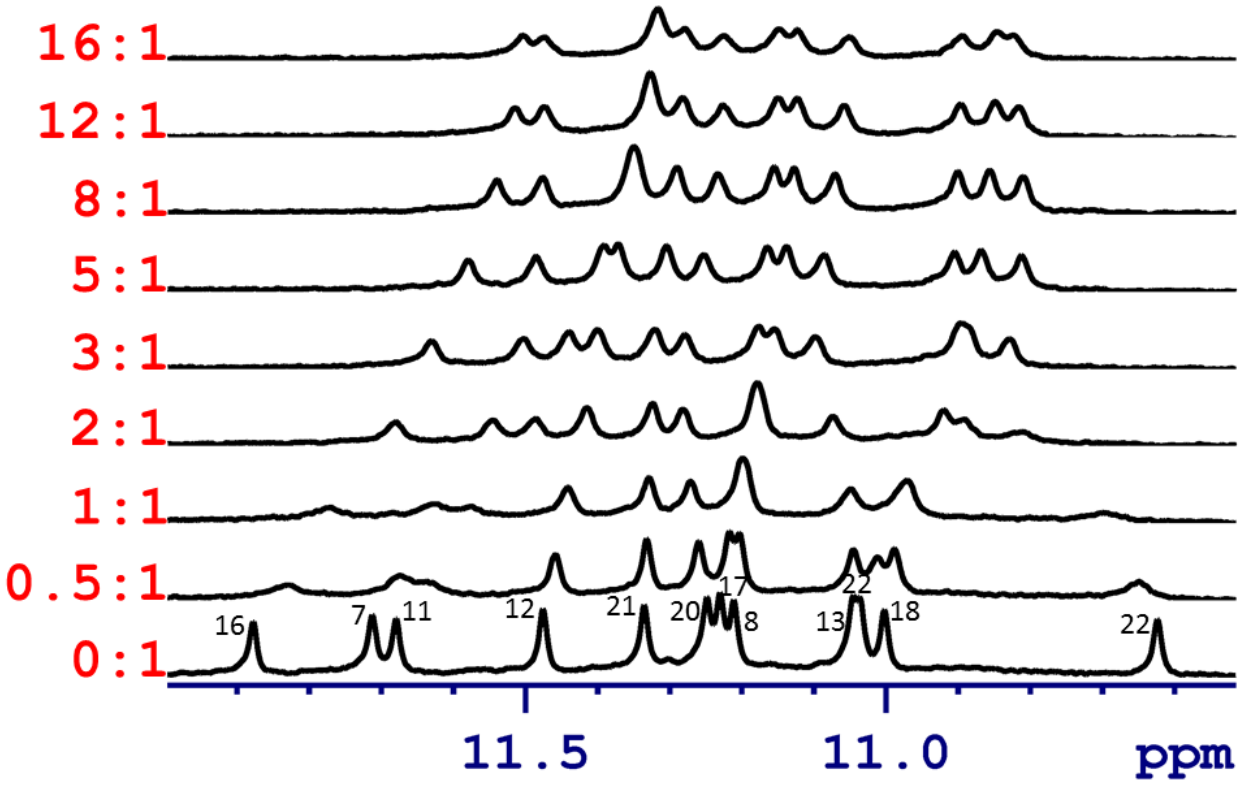


Figure 4: Titration of T-5 with PU-22.
Even at 16:1 ratio, the complex has not yet precipitated out of the solution. Binding occurs on the top and on the bottom tetrad.

4.1.2 EAO-75, 88

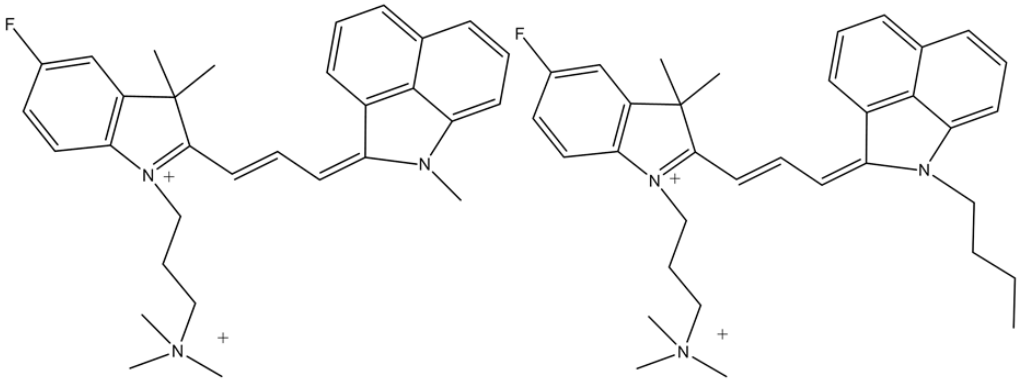
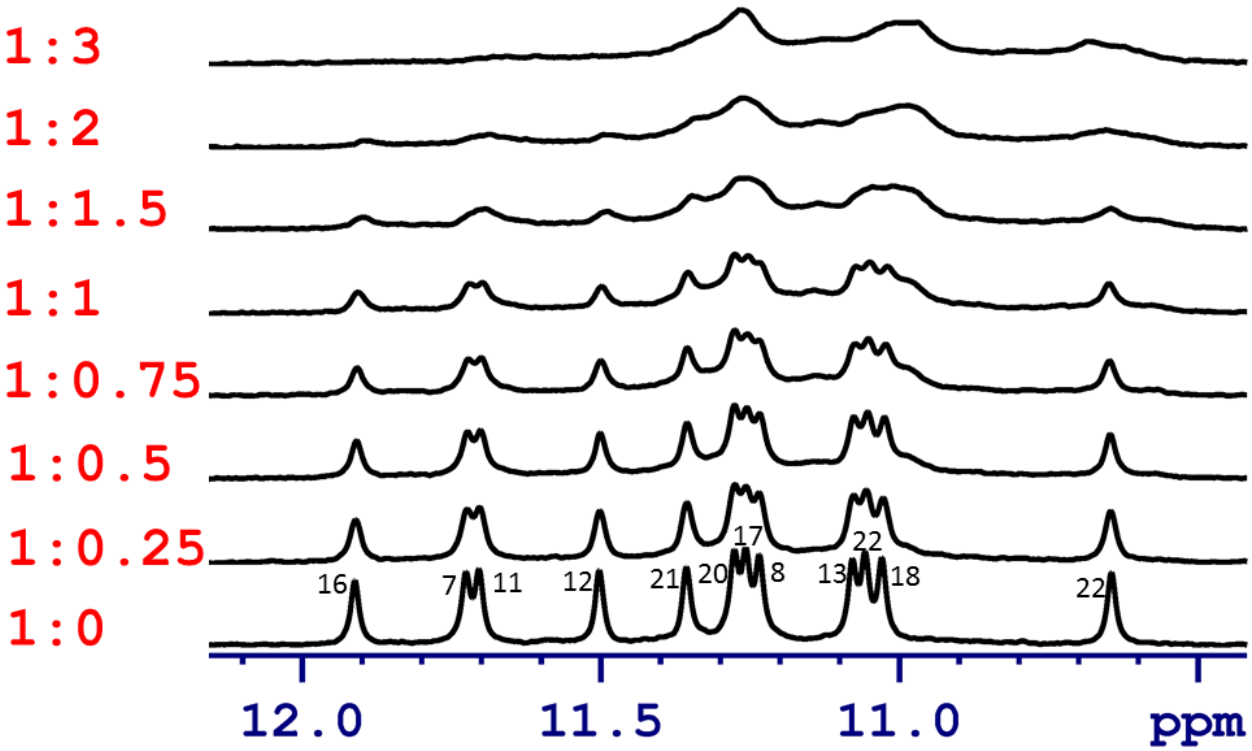


Figure 5: Compounds EAO-75 and EAO-88.
The difference between the two small molecules is in the presence of a longer alkyl chain on the three ring system of EAO-88. Bromine anions (two per molecule) act as counter ions to neutralize any existing positive charges (two per molecule).

4.1.2.1 Titration with PU-22



4.1.2.2 Titration with TEL-24

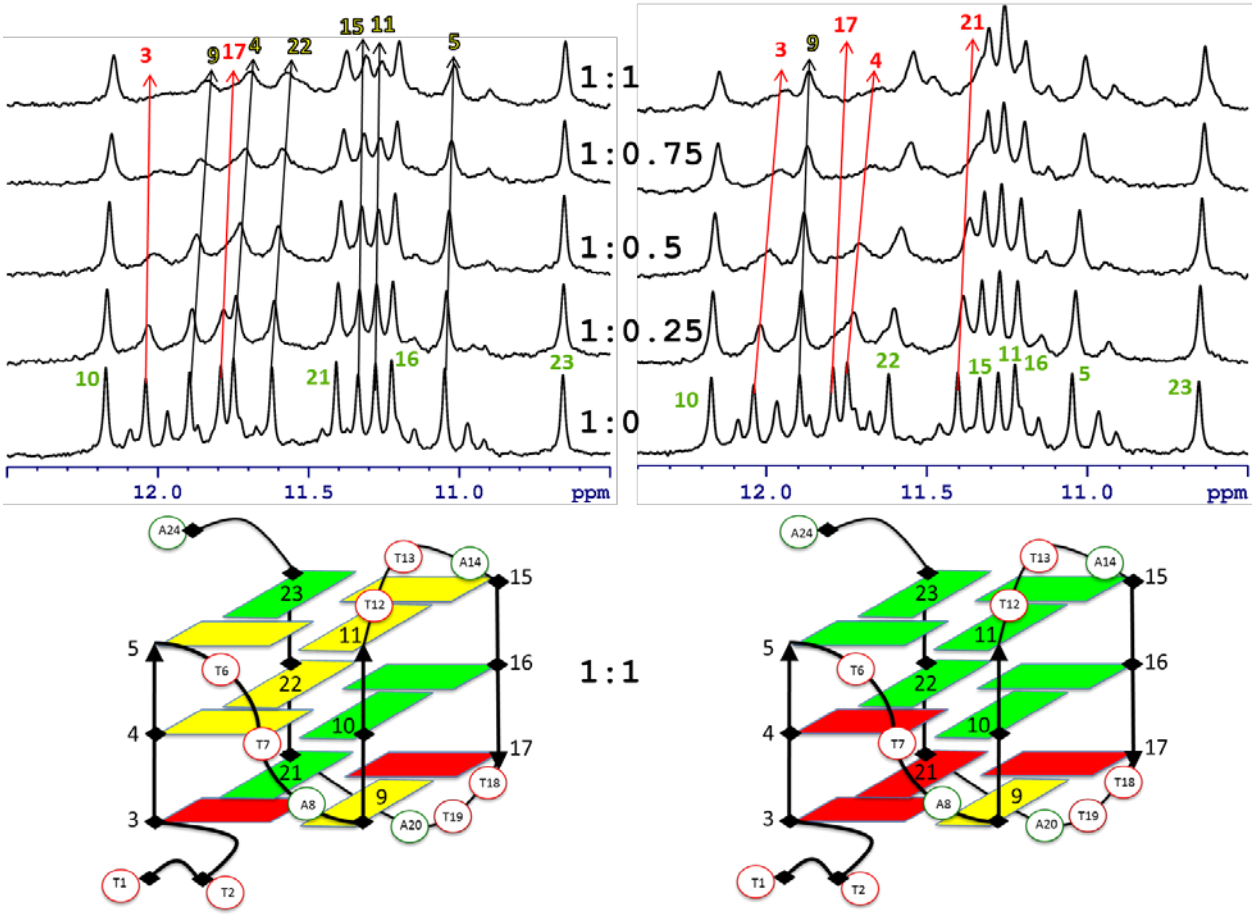


Figure 6: TEL-24 quadruplex titration with EAO-75 (left) and EAO-88 (right) followed with the imino proton spectra of TEL-24.
Ligands were added to the quadruplex at the ratio indicated on the plot, where 1:0 ratio corresponds to the imino proton spectrum of TEL-24 with no ligand added. Schematic structures of TEL-24 are color coded according to the imino protons affected in the titrations. Red residues are strongly affected, yellow are somewhat affected, and green are not affected.